Stochastic Analysis of Energy Networks

This work was financially supported by the Netherlands Organization for Scientific Research (NWO) through the VICI grant 639.033.413.

Stochastic Analysis of Energy Networks

A catalogue record is available from the Eindhoven University of Technology Library
ISBN: 978-90-386-5015-9.

Cover design by Massimiliano Gallo

Printed by Gildeprint Drukkerijen, Enschede

Stochastic Analysis of Energy Networks

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Technische Universiteit Eindhoven, op gezag van de
rector magnificus, prof.dr.ir. F.P.T. Baaijens, voor een
commissie aangewezen door het College voor
Promoties in het openbaar te verdedigen
op donderdag 30 maart 2020 om 16.00 uur

door

Tommaso Nesti

geboren te Pisa, Italïe

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter:	prof.dr. R.W. van der Hofstad
1e promotor:	prof.dr. A.P. Zwart
2e promotor:	dr. A. Zocca (Vrije Universiteit Amsterdam)
leden:	prof.dr.ir. S.C. Borst
	prof.dr. D.T. Crommelin (Universiteit van Amsterdam - CWI)
	prof.dr. J.K. Kok
	prof.dr. J. Moriarty (Queen Mary University of London)
	prof.dr. L. Wehenkel (University of Liège)

Het onderzoek of ontwerp dat in dit proefschrift wordt beschreven is uitgevoerd in overeenstemming met de TU/e Gedragscode Wetenschapsbeoefening.

Acknowledgments

Four and a half years have passed since my arrival in Amsterdam, and many people have played an important role in making this thesis possible. I would like to take this opportunity to express my gratitude to them.

First of all, I wish to thank my advisor Bert Zwart and co-advisor Alessandro Zocca, for their guidance, support and mentoring throughout these years. Bert, you are an extremely driven and passionate researcher, and it is thanks to your direct working style that I was able to get out of my comfort zone and tackle harder challenges in my PhD. Thanks for providing me with so many opportunities to grow as a scientist, such as visiting renowned institutions and doing internships at top companies. Your confidence in my skills played a major role in my decision to seize these opportunities in the first place, and I definitely became more ambitious by working alongside you. Alessandro, your knowledge, patience, and willingness to discuss math problems to the last detail have been invaluable during these years. You always provided me with intelligent insights, regardless of whether my question was about math, academia, or writing style. Most of all, you helped me refocus at times when I was overwhelmed, and I thank you for that.

I would also like to thank the members of my defense committee: Sem Borst, Daan Crommelin, Koen Kok, John Moriarty, and Louis Wehenkel, for taking the time to read this thesis and for providing detailed and insightful comments, and for Remco van der Hofstad for chairing my defense. Special thanks go to John, for envisioning the work that led to Chapter 7 of this thesis, and to Sem, for the many relevant discussions around the results of Chapter 5. I wish to thank Ana Radovanovic, who hosted me as my internship supervisor at Google in 2017 and 2018, and set a powerful example of how it is possible to perform high-level, impactful research in a competitive, fast-paced industry setting. The

results of Chapter 6 of this thesis are based on my internship project with her. Fiona Sloothaak, the results of Chapter 5 have only been possible thanks to your vision, probabilistic skills, and your overall positive and inspiring attitude. Also, thanks for being my travel buddy on several occasions, be it a cold winter school in Germany, or a research trip to sunny California. My institute, CWI, has been a great place to work at. Its atmosphere made it easy for an expat like myself to move to a new country and start a brand new life.

Looking back, I am very grateful to have worked together with several talented researchers. A special mention goes to Jonatha Anselmi, who supervised me during my internship at BCAM while I was still a Master's student. Jonatha, thanks for having introduced me to the world of academic research, and for connecting (and advocating for me!) with the queueing community in Eindhoven, which eventually led me to start my PhD project with Bert. Thanks to my Master's thesis advisor Beatrice Meini, who made it possible for me to work at the intersection of numerical analysis and applied probability, making it easier to transition from Master to Phd.

Izar, Biagio, Simone, Eva and Marcos, you have been the first officemate, colleagues and roommates of my life: here is to Bilbao, where my researcher life set off, and where I left a little bit of hearth. Andre', Lisa, Giada and Irene, you helped me out a lot in my initial transition from Italy to the Netherlands, and my nights out with you are one of the first and most carefree memories about Amsterdam (hard to forget the crazy pirate boat). Chang-Han, our friendship dates back to my first period in Amsterdam, and I hope we'll get to stay in touch. To this day, I am not sure if I am more amazed by your kindness, your mathematical skills, or your huge collection of office-friendly training tools.

Bohan, Paro, Ewan, Bart, Sara, Peter, Brendan, Joost, Daphne, and the rest of the Stochastic group, thanks for having all been part of these four and half intense years of research and climbing. Thanks to all the other CWI colleagues and friends that I crossed my path with, Anton, Marco, Giorgos, Stef, Krzysztof, Deba, you all taught me something unique. Special mention goes to Miha, one of the few persons that can really understand what it means to spend more hours in the office than everywhere else. Thanks for sticking with me in the tough moments, be it a stressful week at work or a heavy(ish) squat. Thanks to Giorgos, for inspiring me to start playing guitar again (I hope we will play together soon).

A proposito, sono passati cinque anni e tocco la chitarra a malapena adesso, ma un abbraccio va ai Sorry 4 Delay. Marco, Bufe, Savio e Luchino, ho un bellissimo ricordo delle nostre prove e serate, la vostra fiducia nelle mie abilità chitarristice ha fatto si che mi mettessi a suonare molto di piu'. Shoutout a Andrea, grande batterista e amico, ci conosciamo da una vita e a distanza di oltre 15 anni continuiamo a beccarci in continenti diversi, stay strong my friend!

Gracias to the artists in my house, Ines and Lucas, for showing me an entirely different side of things. Sometimes we tend to be stuck in our routines, and forget that life is actually more plentiful than we think. And also thanks to my most recent roomies, Ainhoa and Andrea, for bringing a fresh breeze into the house in these last few months.

The italian gang: Macs, Marco, Enrico, Pietro, Ale, Pietro (with the special appearance, directly from Greece, of Dafni!), Giacomo, Nadir, hands down one of the best thing to have happened to me in Amsterdam. Be it our shared love for climbing, music, or just Italian food, your guys transformed my life here.

Macs, thanks for taking me into your house, it meant more than you know. Thanks for all the jam sessions, deep conversations, dinners and training sessions. The amazing cover of this thesis speaks for itself. Pero', impara a fare sto cavolo di nodo a otto che sei inguardabile. Ale, even though we are the same age I have always felt like you were some kind of older brother to me: you were my first climbing teacher (after my failed attempt to introduce you to powerlifting!), and have always been there when I needed to talk, to be calmed down, or to just chill. You have been a vital part of my Amsterdam years, and I hope I was able to give you something back at times when you needed it. Marco, you are one of my oldest friends in Amsterdam. I don't know if it's your sense of humor, or your reserved but confident demeanor, but I always felt I could relate very well to you, and be myself entirely. Enrico "buena onda" Tessadro, I will never fully grasp how you manage to be so positive all the time, but it's infectious! You are always genuinely excited about things, and this shows a wisdom that amazes me.

Thanks to the California buddies, Dave, Joanna, Pablo, Stara, Daniela, Francesca, Yolanda. I arrived there alone, and you guys immediately made me feel welcome. Alle bimbe e ai bimbi di Pisa: Fabio, Annina, Marco, Ele, Borgo, Leo, Massei, Pietro, Ele, Mauro, Sara, Veronica, Tome, I'll never forget our times in Pisa. I feel so lucky that I got to know you, and that we shared so many important moments together. I am so happy every time we get to see each other again.

Nothing (literally) of this would have been possible without the support and love of my family. Babbo, Mamma, grazie per tutta la fiducia che avete riposto in me, e che continuate a riporre giorno dopo giorno. Ceci, grazie di essermi amica, e di essermi restata vicina anche quando ero lontano. Vi voglio bene.

Lucy, these final words have a special meaning for the two of us. Every step we took got us closer to each other, and nobody but us know how many are needed to cover 9000 Km. This is the final one. See you soon!

Contents

Chapter 1

Introduction

Contents

The electricity transmission network is regarded as one of the greatest engineering achievements of the 20th century [46], and is expected to power day-to-day human activities in a reliable and seamless fashion. The increase of intermittent renewable generation such as wind and solar photovoltaics (PV) in the first two decades of the 21st century [155] is making this expectation challenging to live up to.

A well-managed power grid should meet the required power demand at all times, while ensuring that reliability constraints are not violated. Such constraints specify admissible ranges for key quantities, such as transmission line currents and temperatures. For instance, avoiding transmission line overheating is crucial in order to prevent sag and loss of tensile strength, which could result in the tripping of the line.

The system operator traditionally achieves this by making periodic control actions, such as power rescheduling or curtailment, that adapt the operating point of the grid in response to changing conditions. This paradigm relies on the

fundamental assumption that the grid remains roughly *static* between control instants. Such an assumption, while reasonable in the previous century, is hardly realistic for modern grids with ever-growing supply-side uncertainty.

1.1 General background

The inherently *uncertain* nature of renewable energy sources like wind and PV is responsible for significant amounts of variability in power output in the short term [123, 90], caused by changing meteorological conditions. Moreover, the power outputs of individual renewable generators can exhibit considerable correlations. For example, the diurnal cycle of solar power production can result in ramps and shortages of power output from PV generators during different hours of the day [190]. Moreover, the power output from wind farms exhibits correlations which are influenced by the geographical distance between the farms [115], and wind energy follows diurnal cycles as well [108].

Power imbalances caused by generation intermittency may cause grid stability constraints to be violated. For example, 80% of the bottlenecks in the European high-voltage grid in 2015 were already caused by renewables [204]. In order to ensure that stability constraints are not violated, grid operators might reschedule power from flexible dispatchable controllable generators (such as hydroelectric and natural gas power plants) in response to excesses or shortages of renewable generation, and failing that they might resort to curtailing either *power demand* or *power production*. Both measures are undesirable events: on the one hand, modern societies have became used to a continuous and reliable supply of energy; on the other hand, renewable energy is a clean and virtually zero-cost energy source.

In order to meet the objective that 50% of the state's production should come from carbon-free sources by 2030 [36], the state of California has significantly increased its renewable generation capacity in recent years. Unfortunately, curtailments are also on the rise: Fig. 1.1 shows the steady growth of renewable curtailment in the California Independent System Operator (CAISO) in recent years, with historical peak of more than 220,000 megawatt-hours of electricity in May 2019. In the words of CAISO [34]: "The ISO is seeking solutions to avoid or reduce the amount of curtailment of renewable power to maximize the use of clean energy sources."

Furthermore, the increased supply-side uncertainty may result in higher risk of grid components failures, such as transmission lines overloads, which may trigger cascading failures and blackouts, resulting in tremendous economical and societal costs. The frequency of major blackouts increased substantially in recent years [14], with two prominent examples being the Northeast blackout

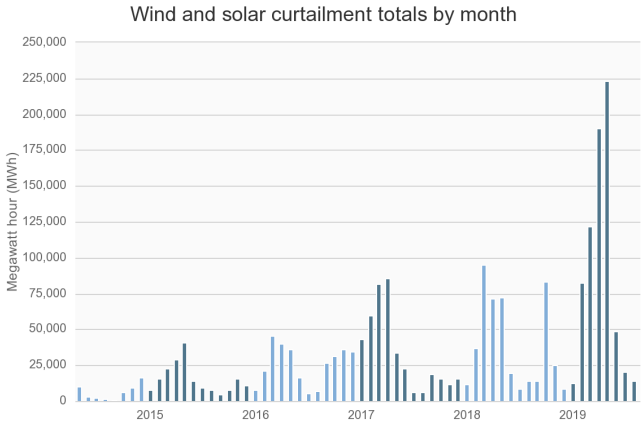


Figure 1.1: Wind and solar power curtailment in California in the period 2014-2019 [34].

in 2003 [183] and the San Diego blackout in 2011 [63].

Energy markets are also affected by the increase in renewable energy production [54, 141]. Energy prices can exhibit significant volatility throughout different hours of the day, and while the exact mechanisms behind energy pricing vary from market to market (see Section 1.3.2 for a description of the Locational Marginal Pricing mechanism, a common market design in the US), a recurrent feature is that prices are usually negatively correlated with the amount of renewable generation in the grid mix [141], as illustrated in Fig. 1.2. In some cases prices can even turn *negative*: Fig. 1.3 illustrates joint occurrences between curtailments of wind generation in the Texas electric grid and real-time negative electricity prices for the West Hub.[1]

In markets adopting the Locational Marginal Pricing (LMP) mechanism, such as the Californian one, prices are location-dependent, and variable supply mix and grid congestion status may result in significant *spatial* price variations, as can be appreciated in Fig. 1.4. In view of the above discussion, it is clear that there is an urgent need to take into account short-term variability in power grid operations.

Quoting the US national academic report [131]: "In short, the greatest achievement of the 20th century needs to be reengineered to meet the needs of the 21st century. Achieving this grid of the future will require effort on

[1] Wind curtailments and corresponding negative prices have substantially dropped after 2011 thanks to Texas's transmission expansions programs [60].

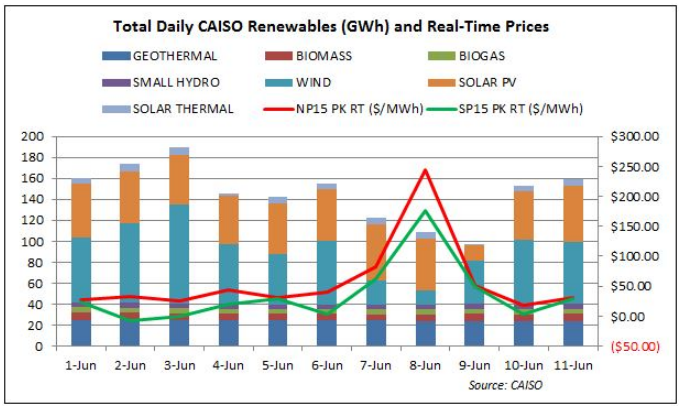


Figure 1.2: Total daily CAISO renewable production and real-time prices, June 2014 [174].

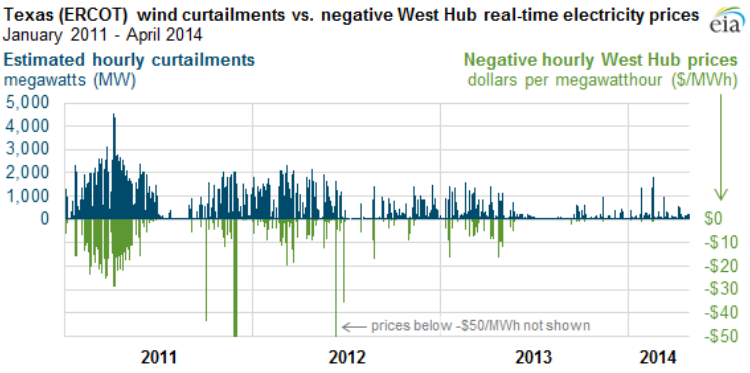


Figure 1.3: Joint occurrences between curtailments of wind generation and real-time negative electricity prices in the Texas electric grid [60].

several fronts. Certainly there is a need for continued shorter-term engineering research and development, building on the existing analytic foundations for the grid. But there is also a need for more fundamental research to expand these analytic foundations."

With this goal in mind, in this thesis we develop novel probabilistic techniques to analyze power grid operations while taking uncertainty into account, which allow us to:

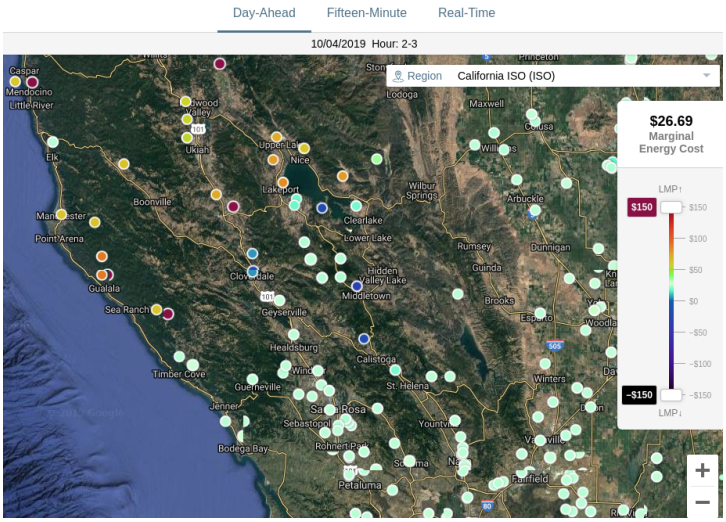


Figure 1.4: CAISO Day-Ahead Locational Marginal Prices on 14th October, 2019, exhibiting significant geographical variability [35].

- Derive probabilistic counterparts of reliability constraints which are analytic and computationally tractable (Chapters 2 and 3).
- Understand microscopic and macroscopic features of cascading failures and blackouts in a probabilistic sense (Chapters 4 and 5).
- Unveil and exploit the relation between energy prices and generation mix in order to forecast price fluctuations (Chapters 6 and 7).

The probabilistic approaches adopted in this thesis include large deviations, concentration inequalities and machine learning techniques, as described in detail in Section 1.3. These techniques offer a very flexible framework that allows us to work with both static and dynamics settings, asymptotic and pre-asymptotic regimes, and microscopic and macroscopic perspectives. Such techniques have been successfully applied in many fields of science and engineering (e.g., communication networks, finance, and queueing theory), but their application to power grid analysis is novel and their potential in this area is mostly unexplored.

This introduction is organized as follows: first, we provide a description of power transmission networks in Sections 1.2 and 1.3, with a particular focus on the optimization paradigm that drives their actual operations, the so-called

Optimal Power Flow problem, and on the corresponding reliability constraints that are of interest to us. The same section also describes the Locational Marginal Pricing mechanism. A review of the existing literature on uncertainty-aware analysis of power grids is provided in Section 1.4. Next, Section 1.5 provides the necessary probabilistic background, while Section 1.6 presents the main contributions of this thesis.

1.2 Power systems modeling

A power grid is an interconnected network for delivering power from producers, or *generators*, to consumers, or *loads*. . The power is transferred over large distances via the *transmission grid*, which delivers power from generators to electrical substations at high voltages, in order to minimize losses. Electrical substations reduce the voltage and inject the power into *distribution systems*, which typically cover a smaller geographical area (such as a city or a neighborhood) and deliver power to individual customers at a lower voltage. This dissertation focuses on the transmission grid, which will be referred to simply as the power grid moving forward, for which we now provide a more detailed description. A power grid can be seen as a connected graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, where nodes $i \in \mathcal{N} = \{1, \dots, n\}$, or *buses*, house (possibly multiple) generators and load, and edges $\ell \in \mathcal{E}$, or *lines*, represent transmission lines that carry power between buses. Let $n = |\mathcal{N}|$ and $m = |\mathcal{E}|$ denote the number of buses and lines, respectively.

1.2.1 Power flow equations

Power is generated and transmitted in alternating current (AC) form, for which we now provide a description (for details, we refer to [14]). Since power can flow in any direction on an edge, the graph that models the power grid is naturally *undirected*. In the following, (k, m) will denote a line joining bus k and m from the perspective of bus k, and (m, k) denotes the same line from the perspective of bus m. For example, the power flowing from bus k to bus m will be denoted by S_{km}, while S_{mk} refers to the power flowing form bus m to bus k. The relation $k \sim m$ means that there is a line between buses k and m, which we denote by $\{k, m\}$ when the orientation is not important.

Let $\mathrm{j} \in \mathbb{C}$ denote the imaginary unit. At time t, let $S_{G_k}(t) = P_{G_k}(t) + \mathrm{j}Q_{G_k}(t) \in \mathbb{C}$ be the complex power produced by bus k, and let $S_{D_k}(t) = P_{D_k(t)} + \mathrm{j}Q_{D_k}(t) \in \mathbb{C}$ be the complex power consumed by bus k. The *net* complex power injected into the grid by bus k is denoted as $S_k(t) = P_k(t) + \mathrm{j}Q_k(t) \in \mathbb{C}$, where $P_k(t) = P_{G_k}(t) + P_{D_k}(t) \in \mathbb{R}$ is known as the *active*

power and $Q_k(t) = Q_{G_k}(t) + Q_{D_k}(t) \in \mathbb{R}$ as the the *reactive power*. Similarly, $S_{km}(t) = P_{km}(t) + \mathrm{j}Q_{km}(t) \in \mathbb{C}$, where $P_{km}(t) \in \mathbb{R}$ is the *active power flow* and $Q_{km}(t) \in \mathbb{R}$ the *reactive power flow*.

If bus k houses a generator but has no load, then $P_k(t) > 0$, whereas if it k has a (positive) load and no generators, $P_k(t) < 0$. If a bus k houses both generators and loads, then the sign of $P_k(t)$ can change according to different demand and generation profiles, with the convention that $P_k(t) > 0$ ($P_k < 0$) means that power is being generated (consumed) at node k.

If bus k is connected to bus m, them bus k injects into transmission line (k, m) a *complex current* $i_{km}(t) \in \mathbb{C}$ at *voltage* $v_k(t) \in \mathbb{C}$, where $v_k(t) = |v_k(t)|e^{\mathrm{j}\delta_k}$, $|v_k(t)| \in \mathbb{R}$ is the *voltage magnitude* and $\delta_k \in (-\pi, \pi]$ is the *voltage phase angle*. For notational simplicity, in the following we suppress the dependence of power, voltage and current on time when not essential. For a complex number x, we denote by x^* its complex conjugate.

The power injected in line (k, m) by bus k satisfies the equation:

$$S_{km} = v_k i_{km}^*. \tag{1.1}$$

The AC power flow equations are governed by two physical laws: Ohm's current law and Kirchoff's current law.

Ohm's current law states that the current flowing on line (k, m) is directly proportional[2] to the voltage drop at these nodes:

$$i_{km}(t) = y_{km}(v_k(t) - v_m(t)), \tag{1.2}$$

where $y_{km} = y_{mk}$ is the (series) *admittance* of line $\{k, m\}$, given by

$$y_{km} = g_{km} + \mathrm{j}b_{km} = \frac{1}{z_{km}} = \frac{1}{r_{km} + \mathrm{j}x_{km}}, \tag{1.3}$$

where the real parameters $g_{km}, b_{km}, z_{km}, r_{km}, x_{km}$ are the *conductance, susceptance, impedance, resistance* and *reactance* of the transmission line, respectively. We use line admittances to build up the *bus admittance* matrix $\mathbf{Y} = \mathbf{G} + i\mathbf{B} \in \mathbb{C}^{n\times n}$ as follows

$$Y_{km} = \begin{cases} -y_{km} & , \quad k \neq m \\ \sum_{l=1}^{n} y_{kl} & k = m \end{cases}, \; G_{km} = \begin{cases} -g_{km} & , \quad k \neq m \\ \sum_{l=1}^{n} g_{kl} & k = m \end{cases}, \; B_{km} = \begin{cases} -b_{km} & , \quad k \neq m \\ \sum_{l=1}^{n} b_{kl} & k = m. \end{cases}$$

In polar form, $Y_{km} = |Y_{km}|e^{\mathrm{j}\alpha_{km}}$. For $\{k, m\} \notin \mathcal{E}$, we set $y_{km} = 0$.

Kirchoff's current law states that the current injection at each bus is equal to the sum of the currents flowing *out* of that bus

$$i_k = \sum_{m\sim k} i_{km} \tag{1.4}$$

[2] Ignoring the effect of shunt components ([14], Chapter 1, Section 1.2).

Combining Kirchoff's law (1.4) and Ohm's law (1.2), we get

$$i_k = \sum_{m \sim k}^{n} y_{km}(v_k - v_m) = \sum_{m=1}^{n} Y_{km} v_m, \tag{1.5}$$

or, in matrix form, $\mathbf{i} = \mathbf{Y}\mathbf{v}$, where $\mathbf{i} = (i_1, \ldots, i_n)^\top \in \mathbb{R}^n$ and $\mathbf{v} = (v_1, \ldots, v_n)^\top \in \mathbb{R}^n$. Combining Eq. (1.1) and Eq. (1.5), we get the AC equations for complex power:

$$S_k = v_k \sum_{m=1}^{n} v_m^* Y_{km}^* = \sum_{m=1}^{n} |v_k||v_m||Y_{km}| e^{\mathrm{j}(\delta_k - \delta_m - \alpha_{km})}. \tag{1.6}$$

In terms of net active and reactive power, the AC equations read

$$P_k = \sum_{m=1}^{n} |v_k||v_m|(B_{km}\sin(\delta_k - \delta_m) + G_{km}\cos(\delta_k - \delta_m)) \tag{1.7}$$

$$Q_k = \sum_{m=1}^{n} |v_k||v_m|(-B_{km}\cos(\delta_k - \delta_m) + G_{km}\sin(\delta_k - \delta_m)), \tag{1.8}$$

and the corresponding power flows are

$$P_{km} = -|v_k|^2 G_{km} + |v_k||v_m|(B_{km}\sin(\delta_k - \delta_m) + G_{km}\cos(\delta_k - \delta_m)), \tag{1.9}$$

$$Q_{km} = |v_k|^2 B_{km} + |v_k||v_m|(-B_{km}\cos(\delta_k - \delta_m) + G_{km}\sin(\delta_k - \delta_m)), \tag{1.10}$$

where we see that $P_k = \sum_{m=1}^{n} P_{km}$, $Q_k = \sum_{m=1}^{n} Q_{km}$. Note that, in general, $P_{km} \neq -P_{mk}$, since $P_{km} + P_{mk} = g_{km}|v_k - v_m|^2 = r_{km}|i_{km}|^2 \geq 0$. The right-hand side of this expression is the amount of active power lost on line $\{k, m\}$ due to resistance.

A classical problem in transmission system analysis consists in determining, given a set of values for generation, demand and voltages, the line power flows that satisfy the the AC equations (1.9). However, the AC equations (1.9) are nonlinear and often analytically intractable, and may not even be well-posed. A solution may not exist for a given set of parameters or, on the other hand, multiple solutions may arise, even for very simple networks [14].

AC equations are usually solved with numerical methods [14], but their non-linearity can introduce significant complexity, making it challenging to routinely solve large-scale Optimal Power Flow problems (see Section 1.3.1) in the normal operational time windows of 5-15 minutes. The efficient calculation of energy prices faces similar challenges, with the additional caveat that the way prices are calculated should be as interpretable and transparent as possible. For this reason, several markets use linearized versions of the power flow equations in order to compute LMPs [33].

1.2.2 DC approximation

In transmission system analysis, the AC power flow equations (1.9) are commonly approximated by a set of simpler linear equations, known as *DC approximation* [151, 150, 175]. Compared to the AC power flow equations, the DC power flow equations are always feasible (provided that the grid is connected), and as such do not suffer from feasibility or multiple solutions issues. The DC approximation stems from three practical observations about high-voltage transmission systems:

- The resistance of the transmission line $\{k, m\}$ is significantly less than the reactance, i.e., $r_{km} \ll x_{km}$. Since $g_{km} = \frac{r_{km}}{r_{km}^2 + x_{km}^2}$ and $b_{km} = \frac{-x_{km}}{r_{km}^2 + x_{km}^2}$, this means that we can approximate $g_{km} \approx 0$ and $b_{km} \approx -\frac{1}{x_{km}}$. Therefore, we assume $Y_{km} = \mathrm{j}B_{km}$, and we define the *weight* of an edge $\{k, m\} \in \mathcal{E}$ as the inverse reactance

$$w_{km} = x_{km}^{-1}. \tag{1.11}$$

  By convention, we set $w_{km} = 0$ if there is no transmission line between k and m, and $w_{km} = w_{mk} > 0$ otherwise.

- Under normal operating conditions, the voltage angle differences $\delta_k - \delta_m$ are small, so we approximate $\sin(\delta_k - \delta_m)$ with $\delta_k - \delta_j$ and $\cos(\delta_k - \delta_m)$ with 1.

- Under normal operating conditions, the voltage magnitudes at the buses are very close to 1 in the *per-unit* system.[3] The DC approximation therefore assumes $|v_k| = 1$ for all buses k.

Finally, the DC approximation only considers active power to describe power flow behavior, ignoring reactive power. Incorporating these simplifying assumptions, Eqs. (1.9) and (1.7) for active power reduce to

$$P_{km} = B_{km}(\delta_k - \delta_m) = -b_{km}(\delta_k - \delta_m) = w_{km}(\delta_k - \delta_m), \tag{1.12}$$

$$P_k = \sum_{m \neq k} P_{km} \tag{1.13}$$

for each bus k and each line (k, m).

In the following, to specify that we are considering active power flows based on the DC approximation, we use the notation $\hat{f}_{km}$ rather than P_{km}, and the

[3] A measuring system that scales all physical quantities by appropriate constants [14], so that resulting values are close to unity.

m-dimensional vector of power flows will be denoted by $\hat{\mathbf{f}} \in \mathbb{R}^m$. Furthermore, since there is no possibility of confusion, we refer to $\hat{\mathbf{f}}$ simply as the vector of power flows, since reactive power flows are not considered under the DC approximation. The DC power flow problem consists in computing the power flows P_{km} given the power injections P_k.

For mathematical convenience, we choose an arbitrary but fixed orientation of the transmission lines, which allows us to denote an edge by the ordered pair $\ell = (i, j) \in \mathcal{E}$. The active power flowing over line (i, j) is denoted by $\hat{f}_{ij}$, with the convention that power is flowing from bus i to bus j if $\hat{f}_{ij} > 0$ and from bus j to bus i if $\hat{f}_{ij} < 0$.

The network topology is described by the *edge-vertex incidence matrix* $\mathbf{A} \in \mathbb{R}^{m \times n}$ defined as

$$A_{\ell,i} = \begin{cases} 1 & \text{if } \ell = (i, j), \\ -1 & \text{if } \ell = (j, i), \\ 0 & \text{otherwise.} \end{cases}$$

Denote by $\mathbf{D} \in \mathbb{R}^{m \times m}$ the diagonal matrix containing the edge weights in Eq. (1.11), defined as $\mathbf{D} = \operatorname{diag}(w_1, \ldots, w_m)$. Finally, the network topology and weights are simultaneously encoded in the *weighted Laplacian matrix* of the graph $\mathcal{G}$, defined as $\mathbf{L} = \mathbf{A}^\top \mathbf{D} \mathbf{A}$ or entry-wise as

$$L_{i,j} = \begin{cases} -w_{ij} & \text{if } i \neq j, \\ \sum_{k \neq j} w_{kj} & \text{if } i = j. \end{cases} \tag{1.14}$$

The matrix $\mathbf{L}$ is symmetric, and if the graph is connected its rows sum up to zero and thus $\mathbf{L}$ is singular with rank $n - 1$. The eigenvalue zero has multiplicity one (thanks to the assumption that the graph $\mathcal{G}$ is connected) and the corresponding eigenvector is $\mathbf{1} = (1, \ldots, 1)^\top$. Denote by $\mathbf{v}_2, \ldots, \mathbf{v}_n$ the remaining eigenvectors of $\mathbf{L}$, which are orthogonal to $\mathbf{1}$ and thus have all zero sum.

Note that Eq. (1.12) reads

$$P_k = \sum_{m \neq k} w_{km}(\delta_k - \delta_m) = (\sum_{m \neq k} w_{km})\delta_k - \sum_{m \neq k} w_{km}\delta_m = \sum_m L_{km}\delta_k,$$

and can thus be rewritten in matrix form as

$$\mathbf{p} = \mathbf{L}\boldsymbol{\delta}, \tag{1.15}$$

where $\mathbf{p}, \boldsymbol{\delta} \in \mathbb{R}^n$ are the vector of net (active) power injections and phase voltage angles, respectively. By adding all the rows of Eq. (1.15) we find that $\sum_{k=1}^n p_k = 0$, implying that power balance between generation and demand must hold at all times. In order to solve the DC power flow problem we need to

solve the linear system $\mathbf{p} = \mathbf{L}\boldsymbol{\delta}$. The matrix $\mathbf{L}$ is singular and $\mathbf{p} \in \mathrm{Im}(\mathbf{L})$, the column spaces of matrix $\mathbf{L}$, therefore such a system has an infinite number of solution spanning a 1- dimensional linear space:

$$\boldsymbol{\delta}(\mathbf{w}) = \mathbf{L}^{+}\mathbf{p} + (\mathbf{I} - \mathbf{L}^{+}\mathbf{L})\mathbf{w}, \quad \mathbf{w} \in \mathbb{R}^{n}, \tag{1.16}$$

where $\mathbf{L}^{+}$ denotes the Moore-Penrose pseudo-inverse of the matrix $\mathbf{L}$. The matrix $\mathbf{L}^{+}$ can be expressed in closed-form as

$$\mathbf{L}^{+} = \left(\mathbf{L} + \frac{1}{n}\mathbf{J}\right)^{-1} - \frac{1}{n}\mathbf{J},$$

where $\mathbf{J} \in \mathbb{R}^{n\times n}$ denotes the matrix with all entries equal to one.

In the literature, there are two common choices for solving Eq. (1.15). The first one reads

$$\boldsymbol{\delta} = \mathbf{L}^{+}\mathbf{p} \tag{1.17}$$

and corresponds to a choice of $\mathbf{w} = \mathbf{0}$. This choice implicitly picks an average value of zero for the nodal voltage phase angles, since $\sum_{k=1}^{n}\delta_k = \sum_{k=1}^{n}\sum_{m=1}^{n}L^{+}_{km}p_m = \sum_{m=1}^{n}p_m\sum_{k=1}^{n}L^{+}_{km} = 0$.

The second commonly used option is to construct a matrix $\bar{\mathbf{L}}$, calculated using the inverse of the $(n-1)\times(n-1)$ sub-matrix obtained from $\mathbf{L}$ by deleting one row and the corresponding column, denoted by $\widetilde{\mathbf{L}}$:

$$\bar{\mathbf{L}} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \widetilde{\mathbf{L}}^{-1} \end{bmatrix}.$$

The standard choice is to delete the first row and the first column. In this case, the first node is used as reference by setting is phase angle δ_1 equal to zero.

We observe that these two procedures, like any other stemming from Eq. (1.16), are equivalent if one is interested in the line power flows, as these latter depend only on the *phase angle differences*, and it can readily be seen from Eq. (1.16) that for every line (k,m) and every pair of $\mathbf{w}, \mathbf{w}' \in \mathbb{R}^{n}$, $\delta_k(\mathbf{w}) - \delta_m(\mathbf{w}) = \delta_k(\mathbf{w}') - \delta_m(\mathbf{w}')$. Choosing the first option, for example, the line power flows (1.12) can be written as a linear transformation of the power injection via

$$\hat{\mathbf{f}} = \widehat{\mathbf{V}}\mathbf{p}, \tag{1.18}$$

where

$$\widehat{\mathbf{V}} = \mathbf{D}\mathbf{A}\mathbf{L}^{+} \in \mathbb{R}^{m\times n} \tag{1.19}$$

is usually referred to as the power transfer distribution factor (PTDF) matrix. Choosing the second option, one the other hand, yields the PTDF

$$\widehat{\mathbf{V}} = \mathbf{D}\mathbf{A}\bar{\mathbf{L}} = [\mathbf{0}\ \mathbf{D}\widetilde{\mathbf{A}}\widetilde{\mathbf{L}}^{-1}] \in \mathbb{R}^{m\times n}, \tag{1.20}$$

where $\widetilde{\mathbf{A}} \in \mathbb{R}^{m\times(n-1)}$ is the matrix obtained by deleting the first columns of $\mathbf{A}$, and $\widetilde{\mathbf{L}}^{(n-1)\times(n-1)}$ the one obtained by deleting the first row and column of $\mathbf{L}$. We remark that, while the matrices defined in Eqs. (1.19), (1.20) are different, they yield the exact same vector of power flows $\mathbf{f}$. In the rest of this thesis, we will use either formulation, depending on the problem at hand.

The (ℓ, k)-th entry of the PTDF matrix quantifies the change in flow on edge ℓ corresponding to a change of power injection at bus k. With each transmission line ℓ is associated the corresponding *line limit* $\bar{f}_\ell > 0$ (Section 1.3.1), which constraints the amount of power that is allowed to flow on it:

$$|\hat{f}_\ell| < \bar{f}_\ell.$$

It is often convenient to express line flows in units of the line limit

$$f_\ell = \hat{f}_\ell / \bar{f}_\ell, \tag{1.21}$$

so that $\mathbf{f}$ is the vector of *normalized power flows*, which can be expressed in vector form as

$$\mathbf{f} = \mathbf{\Lambda}\hat{\mathbf{f}} \in \mathbb{R}^m, \tag{1.22}$$

where $\mathbf{\Lambda}$ is the diagonal matrix $\mathbf{\Lambda} := \mathrm{diag}(1/\bar{f}_1, \dots, 1/\bar{f}_m)$. Correspondingly, the *normalized PTDF* is given by

$$\mathbf{V} = \mathbf{\Lambda}\widehat{\mathbf{V}}. \tag{1.23}$$

1.3 Operations of power systems

Operating a power grid entails addressing multiple design, planning and operational problems, a detailed description of which can be found in the book [14] and is beyond the scope of this thesis. In this section, we describe a simplified version of one of the most important mathematical problems arising in the context of power grid operations.

1.3.1 Optimal Power Flow

The Optimal Power Flow (OPF) problem [94] is an optimization problem that is used to determine the generation schedule that minimize the total system costs while meeting the power demand and satisfying operating constraints of generators and transmission lines. The OPF is run at different time scales, ranging from every 24 hours (for day-ahead planning operations) up to shorter time windows of 5 minutes for real-time operations [16], and it sets generators' output in order to meet the *expected* demand for the upcoming time window. In

its full generality, the OPF is a nonlinear, nonconvex optimization problem, due to the underlying AC power flow equations, which is hard to solve in full generality. These difficulties motivated extensive interest from the power engineering and optimization communities, and a non exhaustive list of solution methods for the general AC OPF include Newton-Raphson methods [190], interior point algorithms, convex relaxations and linearization techniques. For a thorough review of solution techniques, the interested reader is referred to [29].

The inherent difficulties in solving large-scale AC-OPF problems has motivated researchers and practitioners alike to make use of the DC approximation, described in Section 1.2.2. While a simplification of the underlying AC equations, DC-based models are simple and fast, and are commonly used in transmission system analysis [151, 150, 175].

In what follows, we describe a simplified version of the OPF problem based on the DC-approximation, referred to as DC-OPF, which will be used throughout this thesis to demonstrate the potential of the novel mathematical techniques we propose. The DC-OPF is formulated in terms of active power only, and network losses are ignored. For a in-depth discussion on more general AC-OPF formulations, the interested reader is referred to [14]

Let g_k and d_k, respectively, denote the active power produced and consumed at bus k. In the notation of Section 1.2.1, $g_k = P_{G_k}$ and $d_k = P_{D_k}$. We denote the vectors of generation and load as $\mathbf{g} = (g_i)_{i=1}^n \in \mathbb{R}^n, \mathbf{d} = (d_i)_{i=1}^n \in \mathbb{R}^n$, with the convention that if there is no generator (respectively, no load) at node k, we set $g_k = 0$ (respectively, $d_k = 0$). Solving a DC-OPF instance entails determining the generation vector $\mathbf{g}$ (the *decision variable*) that minimizes a convex separable *objective function* of the form $J(\mathbf{g}) = \sum_{i=1}^n J_i(g_i)$, subject to four types of constraints:

- *Power balance:* the generator output $\mathbf{g}$ must meet the *expected* demand $\mathbf{d}$ for the current time window: $\sum_{i=1}^n g_i = \sum_{i=1}^n d_i$.
- *Power flow:* the DC power flow equations (1.12) must be satisfied at all times.
- *Generation:* for each generator k, the amount of power that can be produced is constrained: $\underline{g}_k \leq g_k \leq \bar{g}_k$.
- *Thermal limit of transmission lines:* for each transmission line ℓ, there are constraints on the amount of power that is allowed to flow on it: $|\hat{f}_\ell| < \bar{f}_\ell$. These constraints are particularly important from a reliability perspective because if a line overloads for a sustained period, then it may overheat, sag and lose tensile strength, potentially resulting in the tripping of the line (for instance, by touching the ground or trees). In order to avoid this,

high-voltage transmission lines are endowed with security relays that perform an emergency shutdown as soon as the current flowing in them exceeds a dangerous level, and the limit $\bar{\mathbf{f}}$ is usually set to be lower than the threshold used by the security relays.

Using the notation described in Section 1.2, the DC-OPF problem can be formulated as the following optimization problem:

$$\min_{\mathbf{g}\in\mathbb{R}^n} \quad \sum_{i=1}^{n} J_i(g_i) \tag{1.24}$$

$$\text{s.t.} \quad \sum_{i=1}^{n}(g_i - d_i) = 0 \qquad : \lambda_{\text{en}} \tag{1.25}$$

$$-\underline{\mathbf{f}} \leq \widehat{\mathbf{V}}(\mathbf{g} - \mathbf{d}) \leq \bar{\mathbf{f}} \qquad : \boldsymbol{\mu}^-, \boldsymbol{\mu}^+ \tag{1.26}$$

$$\underline{\mathbf{g}} \leq \mathbf{g} \leq \bar{\mathbf{g}} \qquad : \boldsymbol{\tau}^-, \boldsymbol{\tau}^+ \tag{1.27}$$

$$\tag{1.28}$$

where $J_i(\cdot) : \mathbb{R} \to \mathbb{R}$ denotes the cost function of generation at bus i, which is assumed to be an increasing quadratic function; $\underline{\mathbf{g}}, \bar{\mathbf{g}} \in \mathbb{R}^n$ are the vectors of nodal minimum and maximum generation capacities, respectively; $\underline{\mathbf{f}}, \bar{\mathbf{f}} \in \mathbb{R}^m$ denote the vectors of transmission line limits; $\widehat{\mathbf{V}}$ is the PTDF matrix introduced in Eq. (1.18), and the symbol $\leq$ denotes component-wise inequality. We also denote by $\mathbf{J}(\mathbf{g}) := \sum_{i=1}^{n} J_i(g_i)$ the aggregated cost function.

The variables $\lambda_{\text{en}} \in \mathbb{R}$, $\boldsymbol{\mu}^-, \boldsymbol{\mu}^+ \in \mathbb{R}^m_+$ and $\boldsymbol{\tau}^-, \boldsymbol{\tau}^+ \in \mathbb{R}^n_+$ denote the dual variables of the energy balance constraint (1.25), transmission line constraints (1.26), and generation constraints (1.27), respectively.

The DC-OPF sets an *operating point* (or *base level*) which instructs generators on how much power to produce in the upcoming time window, based on the forecasted load $\mathbf{d}$, which acts as a *parameter* of the problem. We point out that generation and line limits, power grid topology and line reactances are also parameters of the problem, but for our purposes they will be considered fixed over the time scale of interest. [4]

On the other hand, the demand $\mathbf{d}$ does vary on shorter time scales and is thus seen as a *flexible* parameter: it is often of interest to investigate the sensitivity of the optimization problem to a change in the demand parameter, as we do in Chapter 6. The DC-OPF is a strictly convex optimization problem, and, as such, for every demand vector $\mathbf{d}$, there exists a unique optimal solution, which we denote by $\mathbf{g}^* = \mathbf{g}^*(\mathbf{d})$.

[4] We remark that the grid topology can in-fact be altered by means of transmission line switching [79], which are not considered in this thesis.

In practice, real-time loads can deviate from the expected levels, and these deviations are usually handled by Automatic Generation Control (AGC) mechanisms (or, more precisely, by a combination of primary and secondary frequency controls [14]), which operates at smaller time scales than the OPF (seconds to minutes). As these fluctuations are typically small, the scheme based on combining the risk-unaware OPF with automatic real-time adjustments has worked quite well for traditional power grids with low penetration of renewables [16], but as supply-side uncertainty increases a paradigm shift becomes necessary, as we discuss in Section 1.3.3.

1.3.2 Locational Marginal Pricing

Electricity markets designs can exhibit important differences across different parts of the world, reflecting diverse economic and political settings [47]. Locational Marginal Pricing (LMP) is a market architecture adopted by several US markets following the 2003 FERC white paper [64]. Under this architecture, the prices of energy are *nodal* and their calculation is deeply connected with the OPF. Specifically, the LMP at a specific bus is defined as the least cost to service the next increment of demand at that location consistent with all power system operating constraints [140, 112].

Recall that $\mathbf{g}^* = \mathbf{g}^*(\mathbf{d})$ and $\mathbf{J}^* = \mathbf{J}^*(\mathbf{d})$ denote, respectively, the optimal solution and the value function of the DC-OPF in Eqs. (1.24) - (1.27), corresponding to the demand vector $\mathbf{d}$. Denote by $\mathcal{L}$ the Lagrangian function of the DC-OPF, given by

$$\mathcal{L} = \sum_{i=1}^{n} J_i(g_i) - \lambda_{\text{en}} \sum_{i=1}^{n} (g_i - d_i) \tag{1.29}$$

$$- (\boldsymbol{\mu}^+)^\top \left(\bar{\mathbf{f}} - \widehat{\mathbf{V}}(\mathbf{g} - \mathbf{d}) \right) - (\boldsymbol{\mu}^-)^\top \left(\widehat{\mathbf{V}}(\mathbf{g} - \mathbf{d}) - \underline{\mathbf{f}} \right) \tag{1.30}$$

$$- (\boldsymbol{\tau}^+)^\top (\bar{\mathbf{g}} - \mathbf{g}) - (\boldsymbol{\tau}^-)^\top (\mathbf{g} - \underline{\mathbf{g}}). \tag{1.31}$$

In what follows, we give a mathematical definition of LMP based on the simplified DC-OPF formulation described in Section 1.3.1.

Definition 1.1 (LMP [112])**.** *Let $\mathbf{g}^*$ be the unique optimal solution of the DC-OPF in Eqs.* (1.24) - (1.27)*, denote by $\mathbf{J}^*$ the corresponding value of the objective function, and let $\mathcal{L}$ be the corresponding Lagrangian function. The LMP at bus i is the partial derivative of the optimal objective function $\mathbf{J}^*$ with respect to the demand d_i, and is equal to the partial derivative of the Lagrangian $\mathcal{L}$ with respect*

to demand d_i evaluated at the optimal solution: [5]

$$\mathrm{LMP}_i = \frac{\partial \mathbf{J}^*}{\partial d_i} = \frac{\partial \mathcal{L}}{\partial d_i}\Big|_{\mathbf{g}^*}. \tag{1.32}$$

A straightforward calculation (see also [101]) shows that the LMP vector $\mathbf{LMP} = (\mathrm{LMP}_i)_{i=1}^n \in \mathbb{R}^n$ can be represented as

$$\mathbf{LMP} = \lambda_{\mathrm{en}}\mathbf{1} + \widehat{\mathbf{V}}^\top \boldsymbol{\mu} \in \mathbb{R}^n, \tag{1.33}$$

where $\boldsymbol{\mu} = \boldsymbol{\mu}^- - \boldsymbol{\mu}^+ \in \mathbb{R}^m$, and $\mathbf{1} \in \mathbb{R}^n$ denotes a vector of ones.

Definition 1.1 and Eq. (1.33) are based on the DC-OPF formulation, and as such they do not include the effect of active power losses. Most LMP-based markets calculate the LMPs according to Definition 1.1, but they also add a correction accounting for power losses [74]. The loss component is typically negligible [172, 101], and its inclusion goes beyond the scope of this thesis. For a more general discussion of LMPs, which includes a derivation in the case of AC power flow equations, we refer the reader to [112].

We remark that the LMPs, like the dispatched generation $\mathbf{g}$, are an output of the DC-OPF. As every optimization problem, the DC-OPF depends on the parameters that define it. As already discussed, while grid topology and line limits can for most purposes be considered fixed parameters, it is of interest to study the impact of variable parameters, such as the demand $\mathbf{d}$ and uncontrollable renewable generation, on the DC-OPF outputs. In particular, the LMPs depend on changing conditions of nodal demand and uncontrollable generation, and, more in general, on the changing *generation mix* in the grid. This observation is the basis of the work in Chapters 6 and 7.

1.3.3 The role of uncertainty

Modern-day power grids are undergoing a massive transformation, both in terms of decentralization and the introduction of large-scale renewables. Existing transmission grids have been built, for the most part, assuming that generation of electricity is predictable and controllable, and are not designed to accommodate risks caused by large variability.

As mentioned above, the OPF problem computes the most economic dispatch of generation $\mathbf{g}$ that satisfies reliability constraints while meeting the *expected* demand $\mathbf{d}$ for the specific time window. Real-time demand variations are generally small in the time scale of interest of 5-15 minutes [16], and are usually

[5] Eq. (1.32) holds true in the DC-OPF case and, for more general formulations of the OPF, whenever the value function $J^*(\mathbf{d})$ is well-defined and differentiable with respect to $\mathbf{d}$ (cf. envelop theorem [117]).

taken care of by automatic control mechanisms that operate at smaller time scales [14] which do not significantly change the grid operating point. More precisely, the real-time adjustments calculate a modified power output $\mathbf{g}^{\text{rt}}$ that meets the variable real-time demand $\mathbf{d}^{\text{rt}}$. If the load forecasting error $|\mathbf{d} - \mathbf{d}^{\text{rt}}|$ is small, as it is usually the case, the corresponding power injections $\mathbf{g}^{\text{rt}}$ and power flows $\hat{\mathbf{f}}^{\text{rt}} = \widehat{\mathbf{V}}\mathbf{g}^{\text{rt}}$ will be close to the forecast-based OPF outputs $\mathbf{g}, \hat{\mathbf{f}}$, and constraints such as the ones on transmission lines $|\hat{f}_{km}| \leq \bar{f}_{km}$ would be rarely violated [16]. In other words, the frequency control adjustment and load changes are on well-separated time scales. The situation changes when large supply-side uncertainty enters the picture. In the case of real-time significant fluctuations in uncontrollable power output, such as wind and PV generation, the adjusted power output from controllable generators $\mathbf{g}^{\text{rt}}$ may be large, resulting in steep changes in real-time power flows that can violate transmission line constraints.

Even if demand is assumed to be constant, power injections and power flows are effectively *random variables*, and effective reliability analysis of power grids must thus adopt techniques from *probability theory* and *stochastic optimization*.

When a transmission constraint is violated, the corresponding line will be removed form the network (see discussion on thermal limits of transmission lines in Section 1.3.1), causing a global redistribution of power flows which can in turn create stress on the remaining lines. In some cases, this can trigger further outages and result in a *cascading failures* process propagating through the network. The role of supply-side uncertainty is at least two-fold here, one straightforward and the second more subtle. First, the increased risk of *initial* contingencies affects the likelihood of subsequent failures. Second, cascading failures and blackouts in the presence of uncertainty show peculiar features that are observed to a lesser extent (or not at all) in a deterministic setting, as described in Chapter 4 of this thesis. For example, a line failure might occur *endogenously* in the grid as a result of many small correlated fluctuation at individual nodes, and the way failures propagate in the network is often of a non-local nature [99, 103]. Moreover, the complex interaction between the physical structure of the power grid and spatio-temporally correlated random injections results in specific lines having a much higher contingency risk than the majority of the other ones. The traditional deterministic N-1 reliability criterion (Section 1.4.1), in this sense, is insufficient to inform grid operators on the true vulnerabilities of the network: a probabilistic analysis rooted in complex network theory is crucial to understand such phenomena.

Volatile renewable generation is also responsible for considerable fluctuations in electricity prices. As wind and PV energy is significantly cheaper than traditional sources, prices tend to be lower when large amounts of these sources

are available, and can even turn negative (Fig. 1.3). Moreover, steep changes in demand and generation profiles are responsible for different *congestion*[6] of transmission lines. Under the LMP market architecture (Section 1.3.2), the presence of congested lines causes prices to vary wildly across different locations and contribute to their erratic behavior, as depicted in Fig. 1.5.

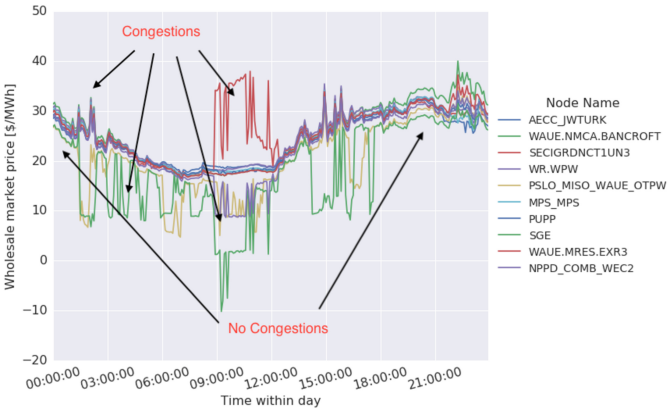


Figure 1.5: Real Time LMPs for randomly selected nodes in the Southern Power Pool (SPP) market [154].

At the same time, understanding and *predicting* fluctuations in electricity prices is highly relevant for wholesale energy market participants, as doing so would allow them to design risk-averse trading strategies while meeting financial and environmental goals.

1.4 Literature Overview

Accounting for short-term variability in power grid operations is of crucial importance, both in terms of the analysis of single contingencies, multiple (cascading) failures, and regarding the impact of supply-side variability on prices. This section presents the relevant literature on the topic.

1.4.1 Probabilistic guarantees on constraint satisfaction

In terms of power grids reliability, a naïve idea would be to come up with a planning that accounts for worst-case behavior of the underlying random

[6] A status describing a line that, while still functioning, reached its operating capacity and no more power is allowed to flow on it.

quantities. Clearly, such an approach is overly conservative and unsuitable for actual operations. This has motivated several recent works that attempt to consider power injection uncertainty by examining various forms of *stochastic guarantees* on constraint violations [197, 191]. This paradigm entails making a planning admissible when the probability of a constraint violation is sufficiently small:

$$\mathbb{P}(\text{constraint violation}) \leq q, \tag{1.34}$$

where $q \in [0, 1]$ is some appropriate threshold. The major difficulty in dealing with chance-constrained optimization is the fact that probabilistic constraints are usually hard, or even impossible, to evaluate analytically. Techniques investigated in the literature include scenario-approach, (rare-event) simulation, robust optimization and chance-constrained optimization. We now describe these approaches in more detail.

The *scenario approach* consists of sampling the relevant uncertain parameter (i.e., wind power production) a certain number of times and use the samples to substitute the chance constraints in Eq. (1.34) with a set of deterministic constraints, resulting in a tractable optimization problem. In [31] the authors provide a bound for the number of samples necessary for the transformed problem to maintain the same level of probabilistic guarantees.

A widely used criterion in traditional deterministic power grid reliability analysis is the so-called N-1 *security criterion* [14]. This rule entails setting the system operating point in such a way that the failure of any single component does not lead to subsequent failures. In [189] the authors formulate a stochastic optimization problem that integrates the (N-1)-based security constraints into a DC-OPF program, and solve using a scenario approach. A similar approach is used in [21] in the context of chance-constrained AC-OPF: the authors set out to solve a chance-constrained AC-OPF via a scenario approach, and tackle the AC power flows' nonlinearities by means of convex approximations. Such methods can be effective but may require a large number of samples. Moreover, modelling stochastic behavior at particular snapshots of time, they are difficult to implement in dynamical continuous-time setting, like the one investigated in Chapter 2 of this thesis.

A related body of literature is based on the idea of handling a chance constraint using simulation techniques, namely (crude) Monte Carlo methods or rare-event simulation. Crude Monte Carlo techniques are based on rewriting a probability $\mathbb{P}(A)$ in the form $\mathbb{E}[\mathbb{1}_A]$, where $\mathbb{1}_A$ denotes the indicator function of the event A. Next, the expectation $\mathbb{E}[\mathbb{1}_A]$ is approximated by the empirical mean taken over repeated independent samples of the underlying random variable.

The scenario-based approach described above is related to this class of methods in that it provides theoretical results on the number of samples necessary

to achieve a desired probabilistic guarantee [30]. Detailed simulations can require a prohibitively large number of samples, especially for events with small probabilities [190], and are thus impractical to be used for short-term operations.

Rare-event simulation aims to handle the prohibitive computational cost of crude Monte Carlo Simulations for estimating very small probabilities, and commonly used techniques are splitting [110, 167], and importance sampling [161]. Rare-event simulation can be faster than crude Monte Carlo methods, but are still not fast enough to use as a subroutine in real-time chance constrained optimization. Moreover, like Monte Carlo methods, they suffer from the lack of analytic expressions that can be useful for gaining theoretical and operational insights in what causes the constraint violations in the first place. Large-deviations techniques offer a solution as they provide a framework to describe the *most likely way* a certain rare event happens.

A *chance-constrained* version of the OPF problem (CC-OPF) is studied in [16], where the authors formulate an OPF with separate chance constraints for each generator and transmission line, and obtain analytic reformulation of the constraints by assuming normally distributed fluctuations ω. For example, for each line (i, j) a probabilistic constraint of the type

$$\mathbb{P}(|f_{ij}(\omega)| > \bar{f}_{ij}) < q$$

is introduced into a DC OPF, where $q \in [0, 1]$ is a small, prescribed threshold, and an affine control scheme is proposed to ensure that supply and demand are matched at all times. The paper [159] generalizes the approach in [16] by considering *weighted chance constraints* (WCC-OPF) which can give different importance to violations of different magnitudes. This entails substituting a constraint of the form

$$\mathbb{P}(y(\omega) > 0) < q,$$

where $y(\omega)$ denotes the *magnitude* of the overload (for example, $y(\omega) = f_{ij}(\omega) - \bar{f}_{ij}$ in the case of upper line limit violation), with

$$\int h(y(\omega))\varphi(\omega)d\omega \leq q,$$

where $\varphi(w)$ is the distribution function of ω, and the *weighting function* $h(y)$, which is non-zero only if $y > 0$, describes the risk related to the overload. WCC-OPF can distinguish between *small* and *large* overloads as it assigns larger weights to larger violations, and as such can be more effective in reducing the probability of large overloads compared to CC-OPF, while being less restrictive when it comes to small violations.

Assuming that ω follows a multivariate Gaussian distribution, the chance-constraints in [16, 158] can be approximated by closed-form expressions, which are not based on large deviations techniques. The large deviations approach considered in this thesis, on the other hand, allows us to develop closed-form approximations for a broader class of stochastic processes, and to deal with more general events such as

$$\mathbb{P}(\exists \text{ line } (i,j) \,:\, |f_{ij}| > \bar{f}_{ij}) < q, \tag{1.35}$$

as described further in Section 1.6.

In [176], the authors propose a computational approach to solve stochastic, *multiperiod* optimal power flow problems based on the *convex relaxation* of a chance-constraints, introduced in [132]: a constraint of the form $\mathbb{P}(f(x,\delta) \leq 0) \geq 1-\alpha$ is rewritten as $\mathbb{E}\psi(t^{-1}f(x,\delta)) \geq \mathbb{P}(f(x,\delta) > 0) \geq 1-\alpha$, where the *generating function* $\psi : \mathbb{R} \to \mathbb{R}$ is a nonnegative, convex function with $\psi(z) > \psi(0) = 1 \; \forall z > 0$.

Most papers in this line of research use the DC power flow equations, and we will make the same assumption throughout this dissertation as well. Recent works on chance-constrained AC-OPF include [21, 188, 157], which handle the AC power flow nonlinearities by means of convex approximations, relaxations and local linearization around a forecasted operating point. A recent work on the analysis of temperature constraints in a discrete-time setting is [102], which focuses on *instantons*, or most-likely events.

1.4.2 Features of cascading failures

In the second part of this thesis, we study features of cascading failures from a probabilistic perspective. To this end, we view the power grid as a *complex network*, where power is produced and consumed at grid nodes, and is transported across the network via transmission lines according to power flow physics. The event of a line failure can cause a global redistribution of power flows, which can in turn create additional strain on the remaining transmission lines. In some cases, this can trigger further outages and result in a *cascading failure* process propagating through the network.

Cascading failure processes are not limited to power grids, and are of interest for the study of many different types of complex networks, including communication networks [107], transport networks [37] and biophysical systems [85]. The microscopic analysis of cascading failures is in general a difficult problem [58, 169, 164], and is particularly challenging in the context of power grids. One of the main reasons is that power flow physics are responsible for a *non-local* propagation of failures across the network, making traditional epidemic models [196, 127, 85, 144] unsuitable for the task. This challenge has

created extensive interest from the engineering and physics communities in the last two decades [1, 2, 3, 49, 50, 124, 129, 128, 80, 105, 162, 179, 206, 196, 201, 199, 200, 205, 48, 92].

Although many different mechanisms for the cascade evolution have been proposed, they all share a common property: the *initial* contingency triggering the cascade is an *external event* that directly leads to the failure of the attacked network component [44, 45, 129]. In view of the discussion in Section 1.4.1 on the impact of supply-side uncertainty on line limits constraint violation, it is important to consider the possibility that line failures can *emerge indirectly* as a result of random, weather-correlated fluctuations of nodal inputs. Such a configuration of nodal inputs is not only the cause of the initial line failure, but can also impact the way subsequent failures propagate in the network. The study of emergent line failures and cascades in power grids is carried out in Chapter 4 of this thesis.

Despite the complexity on a microscopic level that we just discussed, there is empirical evidence for one macroscopic characteristic of blackouts that follows a simple probabilistic law: blackout *sizes* are scale-free [38, 57, 86, 39]. More precisely, if one defines the total size S of a blackout as the number of customers affected, the analysis of historical blackout data [42] reveals that S follows a Pareto distribution, i.e., there exist constants $\alpha > 0, c > 0$ such that, for large x,

$$P(S > x) \approx cx^{-\alpha}, \tag{1.36}$$

where the symbol $\approx$ means that the ratio of both quantities converges to 1 as $x \to \infty$.

Given the tremendous societal impact of large blackouts, understanding the nature of Eq. (1.36) is of major significance. Moreover, the study of power laws is of interest for many other areas of science and engineering, and many possible explanations for their occurrence have been proposed.

In many complex networks applications [10, 144, 73, 153], the scale-free phenomena relate to the scale-free nature of the nodal degree distribution, due to a mechanism known as preferential attachment. However, as the topology of power grids is not scale-free [195], this deduction is not valid in the case of power systems [69].

Another explanation that has been put forward involves the concept of self-organized criticality (SOC) [7], the notion that many systems operate in a critical regime in which many events of interest exhibit power law behavior. In the context of energy networks, it has been suggested that Eq. (1.36) may occur as a consequence of self-organized criticality [38, 57, 14, 178]. These results, which are based on simulations and indirect analogies between blackout models and models known to exhibit SOC, fail to take into account the physics of power

flows and are thus unable to provide a causal explanation for the mechanism behind Eq. (1.36) in power grids.

Other strands of literature models the cascading mechanism as a branching process with critical offspring distribution [104], where each outage (the "parent" in the current generation) triggers a random number of subsequent outages (the "children") with mean equal to one. Such models lead to blackout sizes with infinite mean, corresponding to an exponent $\alpha < 1$ in Eq. (1.36), while recent findings based on advanced statistical analysis of actual blackout data [42] indicate a finite mean blackout size [86, 39] with $\alpha > 1$. For a more complete overview of macroscopic cascading failures models, not limited to power networks, we refer to [170].

Despite the widespread interest that the study of power laws received in the literature, none of the proposed explanations seem adequate to explain the emergence of scale-free blackout sizes in power grids, thus hindering our understanding of network vulnerabilities. In Chapter 5 of this thesis, we propose a different, yet simpler, *causal* mechanism behind the nature of scale-free blackout sizes.

1.4.3 Electricity prices forecasting

As discussed in Section 1.3.3, electricity prices can be highly volatile (see Figs. 1.2 and 1.5). In order to fully understand the cause of such volatility, it is important to realize that the wholesale energy market is different from traditional financial markets, due to the nature of the commodity being traded: electricity cannot be stored in an economically feasible way [198, 100], and it has to be produced and consumed instantly. As presented in Section 1.3.1, supply and demand must be balanced at all times in order to ensure power system stability and, as a consequence, the variable nature of renewable generation and real-time demand is reflected in the volatility of energy prices, in terms of both expected intra-day price variations and sudden, momentary price spikes.

In this thesis, we focus on wholesale energy markets adopting the Locational Marginal Pricing system described in Section 1.3.2. LMPs are used in many US markets, as well as in Singapore, New Zealand and Argentina, while most European markets utilize a *zonal* pricing mechanism [138] (we refer to [67] for a global market overview). One of the main differences between the two mechanism lies in the handling of network congestion. On the one hand, LMPs account for the impact of network congestion in an organic and structured fashion, through the Lagrange multipliers of the OPF. This way, the LMP at a specific node reflects the marginal cost of supplying the next increment of load at that node, consistent with all power grid operating constraints, including transmission line capacities.

Conversely, zonal pricing determines a single price for each zone in the market (which is, in most cases, an entire country) ignoring intra-zonal transmission congestions, which are then managed locally using a variety of different mechanisms. For an overview of the different market structures around the world, and relevant discussion regarding the merits of a potential transition from zonal to nodal pricing in European markets, we refer to [138, 65].

The topic of energy price forecasting has received a lot of attention in the forecasting community in the last 20 years, since the restructuring of energy markets from a government-controlled system to a competitive, deregulated market [198]. In this thesis, we specifically focus on the impact of random renewable generation and demand to prices in LMP-based market, with the goal of developing novel algorithms and techniques to predict price fluctuations. For a study of the impact of uncertainty on prices in European zonal markets, and the development of forecasting techniques, we refer to [141] and references therein, as well as to the review paper [198] for an extensive survey of energy price forecasting techniques.

Thanks to the rich mathematical structure of the LMP mechanism, prediction models for LMPs are not limited to traditional statistical analysis and stochastic models-based techniques, but include structural methods exploiting the mathematical properties of the supply-demand matching process performed by grid operators (namely, the OPF).

The relevant literature on structural prediction models can be categorized based on whether it takes an operator-centric [19, 111, 20, 97] or a participant-centric [101, 209, 70, 18] point of view. In the former case, it is assumed that the modeler has full knowledge of all the parameters defining the OPF formulation, such as generation cost functions, grid topology, and physical properties of the network. Clearly, this allows for the explicit computation of nodal LMPs as dual variables of the corresponding OPF optimization.

In [19], the authors analyze the uncertainty in LMPs with respect to total load in the grid, relying on the structural property that changes in LMPs occur at the so-called critical load levels, under the DC approximation. The calculation of such levels is based on the algorithm proposed in [111]. In [20], the model is extended to the AC power flow framework. In [97] both load and renewable generation uncertainty is considered, and a multiparametric programming formulation that partition the uncertainty space into different critical regions, assuming the DC formulation, is proposed.

The market participant-centric point of view, which relies only on publicly available data without assuming knowledge of the network parameters and operating conditions, has received less attention in the existing literature. The publicly available market data depend on the specific market, and are commonly limited to historical grid-level generation mix (the percentage of total production

from different sources, such as wind or solar), grid-level system loads, and nodal LMPs. On the other hand, OPF parameters such as generation cost functions, generator capacities and transmission line limits are not available.

In [209] the authors utilize the structure of the OPF formulation to infer the congestion status of transmission lines based only on zonal load levels, without considering generation information. Through the concept of System Pattern Regions (SPR), which describe the marginal status[7] of generating units and congestion status of transmission lines, and are based on the notion of critical regions in multiparametric programming [11], zonal prices are obtained by learning the map between zonal load and the corresponding zonal price. In [70] a data-driven approach based on learning nodal prices as a function of nodal loads using support vector machines (SVMs) is presented. The approach is not fully decentralized, since it assumes knowledge of nodal loads, and is computationally unscalable, limiting its applicability to synthetically generated, small grid examples. In [18], the authors assume knowledge of supply bids, nodal generation and nodal prices, and propose an inverse optimization procedure to estimate the remaining parameters of the OPF, which are then used to obtain nodal price predictions. As a result of assuming nodal information on generation, the approach is not fully decentralized and, hence, not suitable for performing predictions from a market-participant perspective.

Finally, in [101] the authors present a methodology to recover grid topology information based only on publicly available historical nodal prices, leveraging results from convex optimization. Although [101] is not concerned with price predictions, the methodology developed therein can play an important role in developing a fully decentralized forecasting algorithm when combined with advanced machine learning techniques and structural properties of the OPF mechanism, as presented in Chapter 6 of this thesis.

The algorithms described above have varying levels of performance, but they all have limitations when predicting extreme price spike values. In the case of price spikes corresponding to rare extreme fluctuations in renewable generation, the worse performance of machine learning-based methodologies can be attributed both to the scarcity of data covering such events, and to the unavoidable errors introduced by a decentralized approach.

Even assuming a centralized perspective, forecasting price spikes is a notoriously difficult problem [113], and is mostly undertaken within the framework of zonal electricity markets. In [78], the authors use a logit model to predict the occurrences of extreme price occurrences using wind power and demand as explanatory variables in the German market. In [28, 143], the authors propose the use of extreme value theory to forecast the extreme tails of electricity

[7] The status of a generator k is *marginal* if $\underline{g}_k < g_k < \bar{g}_k$.

price distributions, with applications to the Nord Pool market and the German energy markets. Regime switching models, which differentiate between a base price regime and higher/lower spike regimes, have been applied to the problem of forecasting extreme prices occurrences in the Australian market [83] and the European Power Exchange EPEX [142, 187]. For a more extensive review of techniques for spike prediction in zonal markets, including stochastic and machine-learning based models, we refer to the literature review in [78].

In Chapter 7 of this thesis, we study the probability of nodal price spikes occurrences in LMP-based markets, a problem that received less attention in the literature compared to zonal markets, proposing an approach which combines the multiparametric programming approach of [11] with large deviations techniques.

1.5 Probabilistic methods

1.5.1 Large deviations results

The theory of large deviations (LD) is concerned with calculating the probabilities of large fluctuations (or *rare events*) that decay exponentially fast as a function of some parameter, such as the magnitude of the noise ε. The theory has found applications in fields such as queueing theory, telecommunication engineering, and finance [27], and its potential for power system applications is investigated for a large part of this thesis. This section provides a brief introduction to the subject.

Consider a sequence of independent, identically distributed random variables $X_1, X_2, \ldots$ with mean μ and variance $\sigma^2 < \infty$, and consider the sequence of sample averages $S_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, $n \geq 1$. Large deviations theory can be seen as a refinement of the Law of Large Numbers and the Central Limit Theorem, as we now show. From the Law of Large Numbers we know that S_n converges in probability and almost surely to the mean μ as $n \to \infty$. In particular, the tail probabilities $\mathbb{P}(|S_n - \mu| \geq \delta)$, for $\delta > 0$, converge to zero as $n \to \infty$. The Central Limit Theorem gives more detailed information about the distribution of S_n, stating that for large enough n the distribution of S_n is close to the normal distribution $\mathcal{N}(\mu, \sigma^2/n)$. In particular, for $\delta > 0$, it holds

$$\mathbb{P}(|S_n - \mu| \geq \delta/\sqrt{n}) \underset{n\to\infty}{\to} \frac{2}{\sqrt{2\pi\sigma^2}} \int_{\delta}^{\infty} e^{-\frac{x^2}{2\sigma^2}} dx. \tag{1.37}$$

The central limit theorem thus deals with deviations of S_n from μ of the order $1/\sqrt{n}$, which are "typical" in the sense that the probability of such a deviation, according to (1.37), is $O(1)$. The theory of large deviations, as the name suggests, deals with *larger* fluctuations. As an illustration, for the particular case of i.i.d.

Gaussian random variables $X_i \sim \mathcal{N}(\mu, \sigma^2)$, the sample average S_n is itself Gaussian with mean μ and variance σ^2/n, so that

$$\mathbb{P}(|S_n - \mu| \geq \delta) = \frac{2}{\sqrt{2\pi\sigma^2}} \int_{\delta\sqrt{n}}^{\infty} e^{-\frac{x^2}{2\sigma^2}} dx \tag{1.38}$$

and therefore

$$\frac{1}{n} \log(\mathbb{P}(|S_n - \mu| \geq \delta)) \underset{n\to\infty}{\to} -\frac{\delta^2}{2\sigma^2}. \tag{1.39}$$

Eq. (1.39) states that the probability that S_n deviates from μ more than δ decays exponentially fast as $e^{-n\delta^2/(2\sigma^2)}$, and is an example of a *large deviation principle*, for which we now give a general definition.

Definition 1.2. *A family of probability measures $\{\mathbb{P}_\varepsilon\}_{\varepsilon>0}$ on a Polish space $\mathcal{X}$ is said to satisfy a large deviation principle (LDP) [55] with rate function I if, for all Borel measurable set $E \subset \mathcal{X}$,*

$$-\inf_{x\in E^\circ} I(x) \leq \liminf_{\varepsilon\to 0} \varepsilon \log\left(\mathbb{P}_\varepsilon(E)\right) \tag{1.40}$$

$$\leq \limsup_{\varepsilon\to 0} \varepsilon \log\left(\mathbb{P}_\varepsilon(E)\right) \leq -\inf_{x\in \bar{E}} I(x), \tag{1.41}$$

where the rate function $I : \mathcal{X} \to [0, \infty]$ is a lower semicontinuous functional on $\mathcal{X}$, and $E^\circ, \bar{E}$ denote, respectively, the interior and the closure of E.

The reason to work with a $\liminf$ and $\limsup$ is mainly technical, and often we can interpret Eq. (1.40) simply as

$$\mathbb{P}_\epsilon(E) \approx \exp(-\inf_{x\in E} I(x)/\varepsilon).$$

The quantity $\inf_{x\in E} I(x)$ is referred to as the *decay rate* of the rare event E, while the term $\arg\inf_{x\in E} I(x)$ is known as its *most likely realization*. The reason for such a name is that every realization x of the rare event E has a certain cost (quantified by the rate function $I(x)$), and the LDP states that the realization with the smallest cost is what dominates in $\mathbb{P}_\epsilon(E)$ as $\varepsilon \to 0$. More precisely, it can be shown (cf. Lemma 4.2 in [68]) that for any neighbourhood B of $\arg\inf_{x\in E} I(x)$, the conditional probability $\mathbb{P}_\varepsilon(\mathcal{X} \setminus B \mid E)$ converges to 0 as $\varepsilon \to 0$.

A very useful tool in the LD arsenal is the *contraction principle*, which allows to map large deviation principles from one space to another [55]. In the context of this thesis, we will often start from a random model of power injections $\mathbf{p}_\varepsilon$, and then use the contraction principle to derive large deviation principles for quantities such as line temperatures and line power flows. For further background, and other engineering applications of LD we refer to [27]; for an introduction of large-deviations theory aimed at physicists, see [182].

1.5.2 Concentration inequalities

Concentration inequalities provide non-asymptotic bounds on the likelihood of a function of many random variables to deviate from its expected value. In contrast to large deviations results, which hold true in the limit as $\varepsilon \to 0$, concentration inequalities are non-asymptotic and are thus valid also in the prelimit. Many concentration bounds have been proved, see [193] for an overview. In our context, we make use of the following result ([193], Theorem 2.26).

Theorem 1.1. *Let $\mathbf{X} = (X_1, \dots, X_n)$ be a vector of i.i.d. Gaussian variables distributed as $\mathcal{N}(0, \sigma^2)$ and let $f : \mathbb{R}^n \to \mathbb{R}$ be L-Lipschitz. Then,*

$$\mathbb{P}\left(|f(\mathbf{X}) - \mathbb{E}(f(\mathbf{X}))| \geq t\right) \leq 2\exp\Big(-\frac{t^2}{2L^2\sigma^2}\Big) \quad \textit{for all } t \geq 0. \tag{1.42}$$

1.6 Contribution of this thesis

This dissertation consists of three main parts. In the first part (Chapters 2 and 3) we focus on deriving chance constrained versions of reliability constraints for single line failures, which are analytic enough to provide insights into strengths and vulnerabilities of the network, and computationally tractable enough to be used for the purposes of short term planning and control. The second part (Chapters 4 and 5) is devoted to understanding the prominent features of multiple line failures and blackouts, both from the microscopic and macroscopic perspective. Finally, the third part (Chapters 6 and 7) focuses on understanding the impact of uncertainty on electricity prices, with the goal of forecasting price fluctuations, uncovering and exploiting the relationship between electricity prices and uncertain renewable production and demand.

Part 1: Uncertainty-aware reliability analysis

In Chapters 2 and 3, we propose techniques based on large deviations theory to handle chance constraints such as in Eq. (1.34). The main idea is to approximate the probability of the constraint violation event E with

$$\mathbb{P}_\varepsilon(E) \approx \exp\Big(-\frac{I^*(E)}{\varepsilon}\Big), \tag{1.43}$$

where ε is "small" and $I^*(E)$ is a particular function of the (rare) event E. Such an approximation is an informal description of a rigorous result known as *large deviation principle*. The parameter ε quantifies the magnitude of the noise in the system. For instance, in the context of stochastic differential equations (SDE) models (Chapter 2), $\sqrt{\varepsilon}$ is a multiplicative constant in front of the Brownian

motion component of the equation, while in the case of static models (Chapters 4 and 7), ε multiplies the covariance matrix of the random vector of interest. In particular, for $\varepsilon = 0$ the system becomes completely deterministic.

The appeal of such a result partially resides in the fact that the term $I^*(E)$, known as *decay rate* of the event E, can often be expressed analytically and in closed form, with obvious benefits from the point of view of scalability of the approach and interpretability of the results. Moreover, large deviation principles can be derived for both random variables and random processes, which allows us to study both static and dynamical frameworks. Finally, the derivation of $I^*(x)$ gives as a byproduct the *most likely way*, or most likely path, for the rare event E to happen, which has important consequences for uncertainty-averse planning. For a rigorous definition of concepts such as large deviation principle, decay rate and most likely path, we refer to Section 1.5.

In Chapter 2, we study the probability of overloading of any transmission line over a given time interval $[0, T]$, and we analytically characterize the *capacity regions* of the grid, i.e., the set of controllable parameters $\boldsymbol{\alpha}$ (such as power injections at time 0) such that this probability stays below a fixed threshold throughout the time window $[0, T]$. That is, we study probabilistic constraints of the form

$$p(\boldsymbol{\alpha}) := \mathbb{P}\Big(\sup_{t\in[0,T]} \max_{(i,j)\in\mathcal{E}} |\Theta_{i,j}(t, \boldsymbol{\alpha})| > \Theta_{ij}^{\max} \Big) < q, \tag{1.44}$$

where $\Theta_{ij}(t, \alpha)$ is the temperature of line (i, j) at time t corresponding to the parameter vector $\boldsymbol{\alpha}$, and we describe capacity regions in the $\boldsymbol{\alpha}$ space of the form

$$\mathcal{R}(q) = \{\boldsymbol{\alpha} \,:\, p(\boldsymbol{\alpha}) < q\}. \tag{1.45}$$

The approach models the stochastic behavior at the process-level by using stochastic differential equations models for random power injections, taking into account the transient relationship between line current and line temperatures. Since line temperature responds gradually to current, a short-lived current overload does not necessarily lead to a temperature overload and as such does not constitute a reliability risk. With this in mind, we develop capacity regions for both current and temperature overloads, and investigate the capacity gains achieved by a less conservative approach.

The derivation is based on Freidlin-Wentzell theory for large deviations, and the regions enjoy convexity properties that make them amenable to be used within OPF in the 5-15 minutes time frame, in contrast with approaches based on long Monte-Carlo simulations. In some particular cases closed-form expression are available, resulting in a chance-constrained OPF version with the same computational complexity as the deterministic counterpart. Even when

closed-form expressions are not available, the existence of efficient algorithms for decay rates calculations [82] makes our approach computationally feasible. Finally, compared to many papers discussed before, which model stochastic behavior at particular snapshots of time, we employ a process-level model that allows to exploit the transient relationships between current and temperature, leading to a less conservative approach. Chapter 2 is based on [133].

The large deviations methods in Chapter 2, while powerful, rely on a scaling procedure and produce results which are theoretically valid in the asymptotic regime $\varepsilon \to 0$. A natural-follow up question is how to develop chance constraints that are valid in the prelimit as well, without any restriction on the magnitude of the noise ε. We set out to this task in Chapter 3, where be derive strict upper bounds for the probability of a line failure $E = \{\max_{\ell \in \mathcal{E}} |f_\ell| \geq 1\}$. Assuming a multivariate Gaussian model for power injections with mean $\boldsymbol{\mu}$ and covariance matrix Σ, we develop bounds of the form

$$\mathbb{P}(E) \leq \phi(\boldsymbol{\mu}, \Sigma),$$

where ϕ is a deterministic function of μ and Σ. Such bounds can be used to derive approximation of chance constraints that are guaranteed to be conservative, and that are explicit enough to be used for optimization purpose on short time scales. This leads to convex polyhedral capacity regions that share similarities with those of Chapter 2, without any assumptions on the magnitude of the noise ε. The mathematical tools behind the derivation of such bounds are concentration inequalities, for which we have provided a short introduction in Section 1.5. Chapter 3 is based on [136].

The methods used in the first part of the thesis, namely large deviations results and concentration inequalities, can be seen as complementary efforts towards the goal of deriving novel chance constrained versions of reliability constraints that can be effortlessly embedded in existing power grid optimization routines.

Compared to [16], a crucial element of the approach taken in this thesis is that we work with chance constraints of the form

$$\mathbb{P}(\exists \text{ line } (i,j) \,:\, |f_{ij}| > \bar{f}_{ij}) < q, \tag{1.46}$$

while in [16] the m constraints

$$\mathbb{P}(|f_{ij}| > \bar{f}_{ij}) < q_{ij} \quad \forall \text{ line } (i,j), \tag{1.47}$$

are used. The constraint in Eq. (1.46) requires that the probability of *any* line overload be smaller than a reliability target q, and thus correctly controls the probability of the actual event that we seek to avoid. Reaching the same reliability level using the approach in Eq. (1.47) would require to use the thresholds

$q_{ij} \leq q/m$, resulting in a much more conservative approach. On the other hand, the constraint in Eq. (1.46) is harder to evaluate analytically than that in Eq. (1.47) but, as it turns out, large deviations and concentration inequalities techniques are powerful enough to handle this difficulty.

Part 2: Features of cascading failures and blackouts

While the first part of this thesis studies the event of single line failures and tries to prevent constraint violations by deriving probabilistic counterparts of reliability constraints to be used in stochastic versions of OPF, the second part (Chapters 4 and 5) focuses on how the world looks like in the case of a line failure. Moreover, it seeks to understand what happens after a failure, in an effort to move beyond the classic N-1 deterministic criterion (described in Section 1.4.1) using a probabilistic framework.

Chapters 4 retains the microscopic perspective adopted in Chapters 2 and 3, and models power grids as complex networks with random, possibly correlated power injections at the nodes, modeling variable renewable production. In contrast to traditional studies on cascading failures, where the initiating event is a deliberate attack (either targeted or random) to the grid stability that triggers subsequent failures (which is also the setting of the N-1 criterion), we study failures that can *emerge endogenously* from the network as a result of random power injections at the nodes, coupled by the network structure and power flow physics. This is a natural follow up study to the research of Chapters 2 and 3, which focused on estimating the probability of line failures and, as it turns out, provides the most likely way for such failures to happen as a byproduct. Here, we deliberately focus on this aspect and use large deviations techniques to explicitly determine the most likely configuration of power inputs leading to line failures, and rank transmission lines according to their failure probability.

Moreover, we are able to predict how subsequent failures will propagate in the network after the first endogenous failure and compare this cascading process to that induced by a purely exogenous disturbance, finding out that cascades can propagate quicker under the novel framework than classical vulnerability analysis. The results are mathematically exact in a small-noise regime, and their accuracy has been validated in a case study using realistic data for the German transmission grid.

Compared to previous studies on cascading failures, which are for the most part based on epidemic models [85, 196], the research presented in Chapter 4 clearly illustrates the potential of using large deviations theory to analyze the novel concept of emergent failures, and its widespread implications for understanding blackouts in complex transport networks under uncertainty. Chapter 4 is based on [137]

The topic of cascading failures in power grids is investigated once more in Chapter 5, from a different perspective. Here, we devote our attention to one well-known macroscopic feature of blackouts: the scale-free nature of blackout sizes. That is, if S denotes the number of customers affected by a blackout, the distribution of S follows a Pareto law, namely there exist constants $C, \alpha > 0$ such that, for large x,

$$P(S > x) \approx Cx^{-\alpha}. \tag{1.48}$$

As discussed in Section 1.4.2, many explanations have been proposed to explain why such a scaling law should emerge, none of which entirely satisfactory.

The main contribution of Chapter 5 is to propose a novel, *causal* and simpler explanation: we argue that the scale-free nature of blackout sizes connects to the scale-free nature of city sizes. It is well-documented that city sizes exhibit a Pareto distribution [38, 57, 86, 39], with a tail index α which is remarkably close to that of blackout sizes. Motivated by this simple observation, we model power grids as graphs with heavy-tailed Pareto-distributed sinks, which represent demand from cities, and study cascading failures on such graphs initiated by an initial disturbance. As power flows get redistributed, more lines can become overloaded and, after a sufficient number of failures, the power grid will eventually disconnect in islanded components. A power shortage (blackout) happens when the generators present in a component are not sufficient to meet the demand, and we show that the distribution of the size of such shortages is determined by that of the cities in the island.

Our approach differs from traditional explanations in that it does not relate scale-free phenomena to the scale-free nature of the network topology, and suggest new ways of approaching such phenomena in other transport networks. Moreover, our analysis shows that in order to make the grid more resilient to large blackouts the focus should be on making the *individual cities* more resilient, as opposed to the commonly suggested approach in the power engineering literature to perform network upgrades. We illustrate how such upgrades do not affect the tail index of the blackout distribution, which is what ultimately determines the likelihood of large blackouts. On the other hand, short-term emergency responses aimed at surviving blackouts of predetermined duration, such as local storage, may prove vital in drastically reducing the economic and societal consequence of large outages, since the distribution of blackout duration has a lighter tail than that of blackout sizes. Chapter 5 is based on [134].

Part 3: Impact of uncertainty on energy prices

The third and final part of this thesis deals with the topic of electricity prices fluctuations. In Chapter 6, we develop a machine learning methodology to predict Locational Marginal Prices (LMP) from a *decentralized perspective*, i.e.,

by using only publicly available data. The topic is particularly important as electricity market price predictions can enable market participants to shape their consumption and supply while meeting their environmental objectives. The decentralized viewpoint if unavoidable when developing market predictions that could be used by market participants, which usually do not have proprietary information on power grid parameters (such as line and generator limits) that are needed to solve the OPF, which is deeply connected to the calculation of LMPs (Section 1.3.2). However, the decentralized perspective makes the analysis much more challenging due to the scarcity of publicly available data, which amounts to aggregated *grid-wide* demand and generation mix (i.e., the fraction of supply coming from various sources such as wind, solar, coal etc), and historical *nodal* prices. On the other hand, ISOs compute LMPs on the basis of complete information on nodal demands, individual generators bounds, transmission line limits, grid topology and physical parameters of the lines, which are all unknown from the market participant perspective.

In order to predict prices in such an under-determined setting, we develop a structured machine learning methodology to recover the salient features of the energy market that copes with scarce, public and high-dimensional market data. Specifically, our methodology is based on (i) exploiting the mathematical properties of the supply-demand matching process to characterize LMPs as deterministic (although unknown) piece-wise affine functions of renewable supply and nodal demand; (ii) using advanced machine learning and convex optimization techniques to infer grid topology and congestion status, and ultimately learning the piece-wise affine function; (iii) predicting prices based on grid-wide load and generation type mix *forecasts*, which acts as surrogate of the corresponding unknown nodal information. Chapter 6 is based on [154].

It turns out that the methodology developed in Chapter 6 performs remarkably well in forecasting intra-day price variations given the restricting assumptions and limited availability of public data, but in some cases fails to correctly predict price spikes, which is a notoriously difficult problem as discussed in Section 1.4.3. For this reason, in Chapter 7 we take the centralized perspective of the grid operator and focus on predicting extreme fluctuations in LMPs. By assuming full knowledge of the power grid parameters, we are able to explicitly derive the deterministic piecewise affine function linking the stochastic input process, modeling renewable generation, which allows us to use large deviations theory to identify the most likely ways for extreme price spikes to happen as a result of unusual volatile renewable generation profiles. This line of work can be seen as a first contribution to the goal of developing a "economically-constrained" stochastic OPF, an augmented version of the OPF which incorporates probabilistic guarantees on the event of extreme price fluctuations. Chapter 7 is based on [135].

Chapter 2

Large Deviations Analysis of Temperature Overloads

Contents

In this chapter we consider the problem of developing tractable probabilistic counterparts for key reliability constraints introduced in Section 1.4.1, such as allowed ranges for current and temperature of a transmission line. In particular, we use large deviations techniques to study the probability of current and temperature overloads in power grids with stochastic power injections, and we characterize the set of admissible nominal power injections (referred to as *capacity regions*) such that the probability of overloading of any line over a given time interval stays below a fixed target.

It was noted in [191] that a transient current overload does not necessarily lead to a temperature overload, due to the fact that line temperature responds gradually to current. In order to investigate this phenomenon, we model power injections at the process-level using stochastic differential equations (SDE), and apply Freidlin-Wentzell theory to derive capacity regions for current and

temperature overloads events. We show how enforcing stochastic constraints on temperature, rather than on current, results in larger capacity regions, and thus in a less conservative approach.

Due to the nonlinear relationship between current and temperature, the decay rate (Section 1.5) for the temperature process is hard to compute explicitly. To address this issue, we derive two tractable approximations: the first is an inner bound, and the second is based on a Taylor series expansion of the decay rate of the temperature overload probability. Both of the two regions coming out of these approximations capture the benefits of incorporating the transient relationship between temperature and current, and they both have the same computational complexity as the current-based capacity region.

Moreover, we prove important convexity properties of the capacity regions, which enable their efficient application for planning and control purposes, such as in OPF formulations. Finally, in the particular case where the random power injections are modeled by an Ornstein-Uhlenbeck (OU) process, we express the capacity regions in closed form.

Chapter outline. The chapter is structured as follows. In Section 2.1, we describe our model for power injections, line currents and line temperatures, based on stochastic differential equations. Sections 2.2 and 2.3 constitute the core of the chapter: using Freidlin-Wentzell theory, we develop and characterize large deviations-based capacity regions for line currents and line temperatures, respectively, and we provide explicit expressions in the particular case that the power injections follow a multivariate OU process. Moreover, we prove important convexity properties of the different regions. In Section 2.4, numerics for the OU case are presented. We summarize and discuss connections to the remaining chapters of this thesis in Section 2.5. Finally, extended proofs are reported in Appendix 2.A.

2.1 System model

2.1.1 Model for the power grid and DC approximation

The network is specified by a connected graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N} = \{0, 1, 2, \cdots, N\}$ is the set of $|\mathcal{N}| = N + 1$ nodes, modelling buses, and $\mathcal{E}$ is the set of $|\mathcal{E}| = m$ edges, representing the transmission lines. After choosing an arbitrary but fixed orientation of the transmission lines, we denote by $\ell = (i, j) \in \mathcal{E}$ the transmission line between buses i and j, and by $x_\ell^{-1} > 0$ the weight of edge $\ell = (i, j)$, corresponding to the inverse *reactance* of that transmission line. By convention, if there is no line between i and j we set the weight to be zero.

Let $\mathbf{p}(t) = (p_i(t))_{i\in\mathcal{N}}$ denote the vector of active net power injections at time t, with the convention that $p_i(t) \geq 0$ ($p_i(t) \leq 0$) means that power is generated (consumed, respectively) at bus i. Node 0 models the slack bus, which ensures that there are no active power imbalances in the network.

Let $\mathbf{I}(t) = (I_\ell(t))_{\ell\in\mathcal{E}}$ be the vector of line currents, and $K(t) = (K_\ell(t))_{\ell\in\mathcal{E}}$ be the vector of line temperatures. Each transmission line ℓ is associated with a thermal limit $K_{\ell,\max}$, which is the maximum permissible temperature of the line [194]. We define $I_{\ell,\max} > 0$ such that if $|I_\ell(t)| = I_{\ell,\max}$ at all times, then $\lim_{t\to\infty} K_\ell(t) = K_{\ell,\max}$. Throughout this chapter, we work with *normalized* currents $\mathbf{Y}(t) = (Y_\ell(t))_{\ell\in\mathcal{E}}$, defined as $Y_\ell(t) = I_\ell(t)/I_{\ell,\max}$. In order to model the relation between power injections and line currents, we make use of the *DC approximation*, described in Section 1.2.2, which leads to a linear relationship between power injections and active power flows $\hat{\mathbf{f}} = \widehat{\mathbf{V}}\mathbf{p}$, where $\hat{\mathbf{f}} = (\hat{f}_1, \ldots, \hat{f}_m) \in \mathbb{R}^m$ and $\hat{f}_\ell$ is the active power flow on line ℓ. Under the DC approximation, one can approximate the line currents with the (active) power flow on the line

$$I_\ell \approx \hat{f}_\ell, \tag{2.1}$$

as noted in [119, 43], so that, in the notation of Section 1.3.1, we have $I_{\ell,\max} = \bar{f}_\ell$.

Recall that the normalized currents are defined as $Y_\ell(t) = I_\ell(t)/I_{\ell,\max}$, and let $\mathbf{\Lambda} = \mathrm{diag}(1/I_{1,\max}, \ldots, 1/I_{m,\max})$. In view of Eq. (1.18) and Eq. (2.1), the active normalized line currents Y can be written as a linear transformation of the power injections $\mathbf{X}$

$$\mathbf{Y}(t) = \overline{\mathbf{V}}\mathbf{p}(t), \tag{2.2}$$

where $\overline{\mathbf{V}} := \mathbf{\Lambda}\widehat{\mathbf{V}}$ and $\widehat{\mathbf{V}} = \mathbf{D}\mathbf{A}\overline{\mathbf{L}}$ is the PTDF matrix defined in Section 1.2.2.

2.1.2 Stochastic and deterministic power injections

We assume that power injections at nodes $1, \ldots, n_w \leq N$ are stochastic, modelling buses housing intermittent renewable power generation. On the other hand, power injections at nodes $n_w + 1, \ldots, N$ are assumed to be deterministic and constant, modeling conventional loads/generators.

We will be interested in capturing the probability of current/temperature overloads over a finite horizon $[0, T]$, which corresponds to the interval between periodic control actions by the grid operator. Thus, the buses in $\{n_w+1, \ldots, N\}$ are those that may be assumed to have a steady power injection over this time scale, denoted by $\boldsymbol{\mu}_D$. Note that the power injection at the slack node 0 is also stochastic, since $p_0(t) = -\sum_{i=1}^n p_i(t)$. The power injection vector is of the form $\mathbf{p}(t) = (p_0(t), \mathbf{X}(t), \boldsymbol{\mu}_D)$, where $\mathbf{X}(t) \in \mathbb{R}^{n_w}$ is the vector of stochastic

injections, and $\boldsymbol{\mu}_D = (\mu_{D,i})_{i=n_w+1}^{N} \in \mathbb{R}^{N-n_w}$. We denote the initial condition for the stochastic power injections by $\boldsymbol{\mu} := \mathbf{X}(0)$, and let $\bar{\boldsymbol{\mu}} := (\boldsymbol{\mu}, \boldsymbol{\mu}_D)$.

In order to make the dependency of the normalized current on stochastic and deterministic power injections more explicit, we note that

$$\mathbf{Y}(t) = \overline{\mathbf{V}}\mathbf{p}(t) = \begin{bmatrix} \mathbf{0} & \mathbf{V} & \mathbf{V_D} \end{bmatrix} \begin{bmatrix} p_0(t) \\ \mathbf{X}(t) \\ \boldsymbol{\mu}_D \end{bmatrix}, \tag{2.3}$$

where $\mathbf{0} = [0, \ldots, 0]^\top \in \mathbb{R}^m$, $\mathbf{V} \in \mathbb{R}^{m \times n_w}$, $\mathbf{V_D} \in \mathbb{R}^{m \times (n-n_w)}$ are the submatrices of $\overline{\mathbf{V}}$ corresponding to stochastic and deterministic injections, respectively. More compactly,

$$\mathbf{Y}(t) = \mathbf{V}\mathbf{X}(t) + \mathbf{y}, \tag{2.4}$$

where $\mathbf{y} := \mathbf{V_D}\boldsymbol{\mu}_D$. We will refer to Eq. (2.4) as the DC power flow equations. The following lemma shows that matrix $\mathbf{V}$ has rank n_w, i.e., the number of stochastic power injections.

Lemma 2.1. *If the network graph is connected, rank* $(\overline{\mathbf{V}}) = N$ *and rank* $(\mathbf{V}) = n_w$. *In particular, the matrix* $\mathbf{V}$ *has linearly independent columns.*

We may interpret $\bar{\boldsymbol{\mu}} = (\boldsymbol{\mu}, \boldsymbol{\mu}_D)$ as the vector of power injections set by the grid operator at time 0 (for example, $\bar{\boldsymbol{\mu}}$ could be the result of an OPF planning based on a forecast for renewable production). Recall that the initial condition for the normalized currents is $\mathbf{Y}(0) = \boldsymbol{\nu}$, where $\boldsymbol{\nu} := \mathbf{V}\boldsymbol{\mu} + \mathbf{y}$. We are interested in scenarios where power grids operate safely, by assuming that the nominal power injections $\bar{\boldsymbol{\mu}}$ are such that the corresponding expected line currents at time $t = 0$ do not exceed the critical level, i.e., $\|\boldsymbol{\nu}\|_\infty = \max_{\ell=1,\ldots,m} |\nu_\ell| < 1$ (possibly, several $|\nu_\ell|$ could be close to their threshold, modeling a high-stress situation). Subsequently, some of the power injections fluctuate randomly because of the variability of the renewable generators. Our focus is to characterize the set of power injection vectors $\bar{\boldsymbol{\mu}}$ such that the probability of current/temperature overload over a finite horizon $[0, T]$ is below a prescribed target p.

2.1.3 Mapping between line current and line temperature

In this section, we describe how line temperature depends on line current. Recall that $K_\ell(t)$ denotes the temperature of line ℓ. We work with *normalized* line temperatures, defined as follows. Let $K_{\text{env},\ell}$ be the ambient temperature around line ℓ. We define the normalized line temperatures $\boldsymbol{\Theta}(t) = (\Theta_\ell(t))_{\ell \in \mathcal{E}}$ as $\Theta_\ell(t) = \frac{K_\ell(t) - K_{\text{env},\ell}}{K_{\max,\ell} - K_{\text{env},\ell}}$. Note that the reliability constraint on line temperatures reads $\|\Theta_\ell\|_\infty < 1$, where $\|f\|_\infty := \max_{t\in[0,T]} \|\mathbf{f}(t)\|_\infty = \max_{t\in[0,T]} \max_{i=1,\ldots,m} |f_i(t)|$ for a continuous function $f : [0, T] \to \mathbb{R}^m$.

In this spirit, in Section 2.3 we characterize the capacity region of the power grid based on bounding the temperature overload probability. In other words, we describe the set of initial power injection vectors $\bar{\boldsymbol{\mu}}$ such that $\mathbb{P}(\|\boldsymbol{\Theta}\|_\infty \geq 1) \leq p$, where p is a prescribed reliability target.

The transient relationship between the normalized temperature Θ_ℓ and the normalized current is given by the ordinary differential equation [147]

$$\tau_\ell \frac{d\Theta_\ell}{dt} + \Theta_\ell = (Y_\ell)^2, \tag{2.5}$$

where $\tau_\ell > 0$ denotes the thermal constant of the transmission line l. Thus, we have

$$\Theta_\ell(t) = \Theta_\ell(0)e^{-t/\tau_\ell} + \frac{1}{\tau_\ell}\int_0^t e^{-(t-s)/\tau_\ell}(Y_\ell(s))^2 ds. \tag{2.6}$$

Note that the instantaneous line temperature depends on the history of the line current process, with an exponentially decaying weight on past values. The parameter τ_ℓ determines the dependence of the instantaneous temperature on past values of current. If τ_ℓ is small, the dependence on past current values becomes weaker, i.e., the line temperature responds more quickly to changes in current. In the limit as $\tau_\ell \downarrow 0$,the response is instantaneous, i.e., $\Theta_\ell(t) = (Y_\ell(t))^2$.

For the sake of simplicity, we assume the initial condition $\Theta_\ell(0) = (Y_\ell(0))^2 = \nu_\ell^2 \quad \forall \ell \in \mathcal{E}$ for line temperatures. Note that ν_ℓ^2 is the steady-state temperature corresponding to a constant line current ν_ℓ.[1] With the above initial condition, let us denote the mapping (2.6) from the current process $\mathbf{Y}$ to the temperature process $\boldsymbol{\Theta}$ as

$$\boldsymbol{\Theta} = \xi_{\boldsymbol{\tau}}(\mathbf{Y}), \tag{2.7}$$

where we emphasize the dependence on the thermal time constants $\boldsymbol{\tau} = (\tau_\ell)_{\ell \in \mathcal{E}}$.

2.1.4 Stochastic model for power injections

We now describe our stochastic model for the power injections $\mathbf{X}(t)$. Recall that in order to characterize the capacity region of the power grid, we have to estimate the following overload probabilities:

$$\mathbb{P}(\|\mathbf{Y}\|_\infty \geq 1), \quad \mathbb{P}(\|\boldsymbol{\Theta}\|_\infty \geq 1).$$

[1] This assumption, which ignores the history of the temperature process prior to time $t = 0$, is a natural engineering assumption if the past line temperatures are unavailable. If such measurements are available, it is possible to incorporate these into our capacity region based on temperature overload (Section 2.3.1) as well as its inner bound (Section 2.3.2), although the analysis gets more complicated (see [192, Section 4.3]).

We model the random power input sources as small-noise stochastic differential equations (SDE), for which a comprehensive and sufficiently explicit theory of large deviations is available. SDEs are a flexible modeling tool for continuously varying processes, and their use for wind speed modeling has been adopted recently by several authors [192, 95, 125, 96]. Formally, we model the vector of random power injections $\mathbf{X}^\varepsilon(t) = (X_1^\varepsilon(t), \ldots, X_{n_w}^\varepsilon(t))$ as the strong solution of the n_w-dimensional stochastic differential equation (SDE)

$$d\mathbf{X}^\varepsilon(t) = \mathbf{b}(\mathbf{X}^\varepsilon(t))dt + \sqrt{\varepsilon}\,\mathbf{\Gamma}(\mathbf{X}^\varepsilon(t))d\mathbf{W}(t), \quad t \geq 0, \tag{2.8}$$

where $\mathbf{X}^\varepsilon(0) = \boldsymbol{\mu}$, $\mathbf{b}(x) = (b_1(x_1), \ldots, b_{n_w}(x_{n_w}))$, $\mathbf{\Gamma}(x) = \text{diag}(\{\gamma_i(x_i)\}_{i=1}^{n_w})$ and $\mathbf{W}(t) = (W_i(t))_{i=1,\ldots n_w}$. The function $\mathbf{b}$ is referred to as the *drift function*, and captures the evolution of the process in the absence of noise. The noise in the evolution of the process is introduced by the second term in Eq. (2.8): $W_i(t)$ is a standard Brownian motion in $\mathbb{R}$. This noise is modulated in a state-dependent fashion by the *diffusion function* $\mathbf{\Gamma}$, and the scaling parameter $\varepsilon > 0$ captures the amount of randomness in the power injections. As $\varepsilon \to 0$, the magnitude of the noise injected into the evolution of the process $\mathbf{X}^\varepsilon(t)$ diminishes, making large deviations from the "noise-less" behavior exponentially (in $1/\varepsilon$) unlikely.

It is in this regime that LD theory gives us tractable approximations of the probabilities of the rare events corresponding to current and temperature overloads. In practice, ε can be chosen so that the variance of the process $\mathbf{X}^\varepsilon(t)$ matches the estimation error for renewable production over a specific time unit (Section 2.4.2). We make the following regularity assumptions: $\forall i = 1, \ldots, m$, $b_i : \mathbb{R} \to \mathbb{R}$ is Lipschitz continuous and differentiable with $b_i(\mu_i) = 0$; $\gamma_i : \mathbb{R} \to (0, \infty)$ is Lipschitz continuous, bounded and differentiable.

The $\varepsilon-$scaled current process $\mathbf{Y}^\varepsilon(t) = (\mathbf{Y}^\varepsilon(t))_{\ell \in \mathcal{E}}$ is defined as per the DC power flow equations: $\mathbf{Y}^\varepsilon(t) = \mathbf{V}\mathbf{X}^\varepsilon(t) + \mathbf{y}$. Similarly, the $\varepsilon-$scaled temperature process $\mathbf{\Theta}_{\boldsymbol{\tau}}^\varepsilon(t) = (\Theta_{\boldsymbol{\tau}}^\varepsilon(t))_{\ell \in \mathcal{E}}$, with thermal constant $\boldsymbol{\tau}$, is defined as $\mathbf{\Theta}_{\boldsymbol{\tau}}^\varepsilon = \xi_{\boldsymbol{\tau}}(\mathbf{Y}^\varepsilon)$, where the map $\xi_{\boldsymbol{\tau}}$ is given in Eqs. (2.6) - (2.7). In the following sections, we apply the theory of large deviations to estimate the probabilities $\mathbb{P}(\|\mathbf{Y}^\varepsilon\|_\infty \geq 1)$ and $\mathbb{P}(\|\mathbf{\Theta}^\varepsilon\|_\infty \geq 1)$, in the limit as $\varepsilon \downarrow 0$.

2.2 Capacity regions based on current overload

The traditional approach for ensuring line reliability is to impose the condition $\|\mathbf{Y}(t)\|_\infty := \max_{\ell \in \mathcal{E}} |Y_\ell(t)| < 1$ at all times. In this spirit, in this section we characterize the capacity region of the power grid obtained by bounding the probability of current overload over $[0, T]$ by a prescribed target q

$$\mathbb{P}(\|\mathbf{Y}\|_\infty \geq 1) \leq q.$$

Our focus is to characterize the space of initial power injections that can be 'set' at time 0, such that the probability that the inherent variability in the stochastic sources leads to a current overload before the next control instant is small.[2] The above approach is in line with the conventional technique of enforcing the thermal limits of transmission lines by capping the peak current on each line. In Section 2.3.1 a more refined approach, taking into account the transient relationship between line current and line temperature, is presented.

In the following, we first provide a large deviation principle for the current overflow event $\{\|\mathbf{Y}^\varepsilon\|_\infty \geq 1\}$ in the limit as $\varepsilon \downarrow 0$. Next, we use this characterization to define the current-overload based capacity region, and prove a convexity result which facilitates its application as a constraint in OPF. We then provide two lemmas that are useful for computing the capacity region in practice and we give a closed-form characterization of the capacity region when the stochastic injections follow an OU process.

2.2.1 Large deviations results

As described in Section 1.5.1, the theory of large deviations (LD) is concerned with calculating the exponential decay of rare events probabilities, by means of the so-called *rate functions*. This subsection is based on the Freidlin-Wentzell theory [55], which is concerned with large deviation principles for the *paths* of a stochastic process. Thanks to Theorem 5.6.7 in [55], the power injections process $\mathbf{X}^\varepsilon$ satisfies a sample path large deviation principle (SPLDP) over the function space $C_{\boldsymbol{\mu}}([0,T]) = \{\mathbf{g} : [0,T] \to \mathbb{R}^{n_w} : \mathbf{g} \text{ is continuous and } \mathbf{g}(0) = \boldsymbol{\mu}\}$, with good rate function

$$\mathcal{I}_{\text{pow}}(\mathbf{g}) = \sum_{i=1}^{n_w} \mathcal{I}_{\text{pow},i}(g_i). \tag{2.9}$$

Here, $\mathbf{g} = (g_1, \ldots, g_{n_w})$ and $\mathcal{I}_{\text{pow},i}$ is the good rate function for the SPLDP associated with the process $X_i^\varepsilon(t)$, $i = 1, \ldots, n_w$, and it is given by

$$\mathcal{I}_{\text{pow},i}(g_i) = \begin{cases} \frac{1}{2}\int_0^T \left(\frac{g_i' - b_i(g_i)}{\gamma_i(g_i)}\right)^2 dt & \text{if } g_i \in H^1_{\mu_i}(\mathbb{R}), \\ \infty & \text{if } g_i \notin H^1_{\mu_i}(\mathbb{R}). \end{cases}$$

Here, $H^1_{\boldsymbol{\mu}}(\mathbb{R}^{n_w}) := \{\mathbf{g} : [0,T] \to \mathbb{R}^{n_w} : \mathbf{g}(t) = \boldsymbol{\mu} + \int_0^t \boldsymbol{\phi}^{(\ell)}(s)ds, \boldsymbol{\phi}^{(\ell)} \in L_2([0,T])\}$ is the space of absolutely continuous functions with value $\boldsymbol{\mu}$ at time $t = 0$ and which possess a square integrable derivative. Next, we apply a

[2] Given the equivalence between line currents and power flows under the DC approximation, the results in this section can also be interpreted in terms of the probability of exceeding line power flow limits.

very useful tool in LD theory, known as the *Contraction Principle*, which allows to map large deviations principles from one space to another. Thanks to the Contraction Principle, Theorem 4.2.1 in [55] and Eq. (2.4), the current process $\mathbf{Y}^\varepsilon$ satisfies a SPLDP with good rate function

$$\mathcal{I}_{\text{cur}}(\mathbf{f}) = \inf_{\substack{\mathbf{g}\in H^1_{\boldsymbol{\mu}}:\\ \mathbf{y}+\mathbf{V}\mathbf{g}=\mathbf{f}}} \mathcal{I}_{\text{pow}}(\mathbf{g}).$$

Thanks to Lemma 2.1, the matrix $\mathbf{V}$ has linear independent columns. Therefore, its Moore-Penrose inverse has an explicit formula $\mathbf{V}^+ = (\mathbf{V}^\top\mathbf{V})^{-1}\mathbf{V}^\top$ and it is a left inverse of $\mathbf{V}$. Thus, for $\mathbf{f} \in \mathbf{y} + \mathbf{V}(H^1_{\boldsymbol{\mu}}(\mathbb{R}^{n_w})) \subset H^1_{\boldsymbol{\nu}}(\mathbb{R}^m)$ the equation $\mathbf{y} + \mathbf{V}\mathbf{g} = \mathbf{f}$ has unique solution $\mathbf{g} = \mathbf{V}^+(\mathbf{f} - \mathbf{y})$, yielding

$$\mathcal{I}_{\text{cur}}(\mathbf{f}) = \begin{cases} \mathcal{I}_{\text{pow}}(\mathbf{V}^+(\mathbf{f}-\mathbf{y})) & \text{if } \mathbf{f} \in \mathbf{y} + \mathbf{V}(H^1_{\boldsymbol{\mu}}(\mathbb{R}^{n_w})), \\ \infty & \text{otherwise.} \end{cases} \tag{2.10}$$

For the current overload event we then have that

$$\limsup_{\varepsilon\to 0} \varepsilon \log \mathbb{P}(\|\mathbf{Y}^\varepsilon\|_\infty \geq 1) = -\mathcal{I}^*_{\text{cur}}, \tag{2.11}$$

$$\mathcal{I}^*_{\text{cur}} = \inf_{\substack{\mathbf{f}\in\mathbf{y}+\mathbf{V}H^1_{\boldsymbol{\mu}}:\\ \|\mathbf{f}\|_\infty\geq 1}} \mathcal{I}_{\text{cur}}(\mathbf{f}) = \inf_{\substack{\mathbf{g}\in H^1_{\boldsymbol{\mu}}:\\ \|\mathbf{y}+\mathbf{V}\mathbf{g}\|_\infty\geq 1}} \mathcal{I}_{\text{pow}}(\mathbf{g}), \tag{2.12}$$

with $\mathcal{I}^*_{\text{cur}}$ the *decay rate* for the current overload event.[3]

2.2.2 Derivation of capacity region

Eq. (2.11) yields the following approximation for the current overload probability for small ε:

$$\mathbb{P}(\|\mathbf{Y}^\varepsilon\|_\infty \geq 1) \approx e^{-\mathcal{I}^*_{\text{cur}}(\overline{\boldsymbol{\mu}})/\varepsilon}. \tag{2.13}$$

We use the above approximation to define the capacity region for the power grid, based on the constraint that the probability of current overflow must not exceed p, where $p > 0$ is a small pre-defined threshold:

$$\tilde{\mathcal{R}}^{(\text{cur})}_{\varepsilon,p} := \{\overline{\boldsymbol{\mu}} \in \mathbb{R}^N \,:\, \mathcal{I}^*_{\text{cur}}(\overline{\boldsymbol{\mu}}) \geq -\varepsilon\log(p)\}. \tag{2.14}$$

In the remainder of this section, we shed light on structural properties and computational aspects of this capacity region. Our first result shows that the capacity region is convex with respect to the deterministic power injections.

[3] Note that $\mathbf{f} \in H^1_{\boldsymbol{\nu}} \setminus (\mathbf{y} + \mathbf{V}H^1_{\boldsymbol{\mu}})$ implies $\mathcal{I}_{\text{cur}}(\mathbf{f}) = \infty$, thus $\mathcal{I}^*_{\text{cur}} = \inf_{\mathbf{f}\in\mathbf{y}+\mathbf{V}H^1_{\boldsymbol{\mu}}:\,\|\mathbf{f}\|_\infty\geq 1} \mathcal{I}_{\text{cur}}(\mathbf{f}) = \inf_{\mathbf{f}\in H^1_{\boldsymbol{\nu}}:\,\|\mathbf{f}\|_\infty\geq 1} \mathcal{I}_{\text{cur}}(\mathbf{f})$.

Lemma 2.2. *$\tilde{\mathcal{R}}_{\varepsilon,p}^{(cur)}$ is convex in the deterministic power injections vector $\boldsymbol{\mu}_D$.*

Lemma 2.2 is important as convexity enables the set of allowable deterministic injections to be incorporated as a constraint in OPF problems (see, for example, [16]). For the special case where power injections are modeled as an OU process, we show in Section 2.2.3 that the capacity region $\tilde{\mathcal{R}}_{\varepsilon,p}^{(\text{cur})}$ itself is convex. Letting

$$\psi_\ell = \inf_{\substack{\mathbf{f}\in H^1_{\boldsymbol{\nu}}:\\ \|f_\ell\|_\infty \geq 1}} \mathcal{I}_{\text{cur}}(\mathbf{f}) = \inf_{\substack{\mathbf{g}\in H^1_{\boldsymbol{\mu}}:\\ \|y_\ell + V_\ell \mathbf{g}\|_\infty \geq 1}} \mathcal{I}_{\text{pow}}(\mathbf{g}),$$

with V_ℓ being the ℓ-th row of matrix $\mathbf{V}$, we note that $\mathcal{I}^*_{\text{cur}} = \min_{\ell\in\mathcal{E}'} \psi_\ell$, where $\mathcal{E}' := \{\ell \in \mathcal{E} \,:\, V_\ell \neq 0\}$. [4]

In other words, the decay rate for a current overload in the network is the minimum of the decay rates corresponding to the overload of each line. Decay rates, together with Eq. (2.13), provide an analytical tool to rank transmission lines in terms of their vulnerability. The next lemma shows that the current overload on any line most likely occurs at the end time.

Lemma 2.3. $\forall\, \ell \in \mathcal{E}'$, $\psi_\ell = \inf_{\mathbf{g}\in H^1_{\boldsymbol{\mu}}:\, |y_\ell + V_\ell \mathbf{g}(T)|=1} \mathcal{I}_{pow}(\mathbf{g})$.

For $a \neq \nu_\ell$, define

$$\psi_\ell^{(a)} = \inf_{\mathbf{f}\in \mathbf{y}+\mathbf{V}H^1_{\boldsymbol{\mu}}:\, f_\ell(T)=a} \mathcal{I}_{\text{pow}}(\mathbf{f}), \tag{2.15}$$

so that $\psi_\ell = \psi_\ell^{(1)} \wedge \psi_\ell^{(-1)} = \min \psi_\ell^{(1)}, \psi_\ell^{(-1)}$ and

$$\mathcal{I}^*_{\text{cur}} = \min_{\ell\in\mathcal{E}'} \psi_\ell^{(1)} \wedge \psi_\ell^{(-1)}. \tag{2.16}$$

Eq. (2.16) allows us to rewrite the capacity region as

$$\tilde{\mathcal{R}}_{\varepsilon,p}^{(\text{cur})} = \bigcap_{\ell\in\mathcal{E}',\, a\in\{-1,1\}} \{\overline{\boldsymbol{\mu}} \in \mathbb{R}^N \,:\, \psi_\ell^{(a)} \geq -\varepsilon \log(p)\}. \tag{2.17}$$

Thus, obtaining the capacity region $\tilde{\mathcal{R}}_{\varepsilon,p}^{(\text{cur})}$ hinges on computing $\psi_\ell^{(a)}$, which by definition is the solution of Eq. (2.15). To solve this variational problem with boundary constraints, one can for instance use the Euler - Lagrange equations (see also our discussion in Section 2.4). For simple diffusion models, this approach can be used to obtain the optimal path and $\psi_\ell^{(a)}$ in closed form, leading to an explicit characterization of the capacity region $\tilde{\mathcal{R}}_{\varepsilon,p}^{(\text{cur})}$. Next, we illustrate this for the case where the power injections are modeled as an OU process.

[4] Note that if $V_\ell = 0$, then $Y_\ell^\varepsilon(t) = Y_\ell^\varepsilon(0) = y_\ell$ is constant and $|y_\ell| = |\nu_\ell| < 1$, yielding $\psi_\ell = \inf_{\mathbf{g}\in H^1_{\boldsymbol{\mu}}:\, \|y_\ell\|_\infty \geq 1} \mathcal{I}_{\text{pow}}(\mathbf{g}) = \infty$.

2.2.3 Explicit computations for Ornstein-Uhlenbeck process

In this section we suppose that the power injections $\mathbf{X}^\varepsilon(t)$ follow a multivariate Ornstein-Uhlenbeck (OU) process, which is the most tractable example of an SDE and is, in particular, a Gaussian process.[5] Such a process is of the form

$$d\mathbf{X}^\varepsilon(t) = \mathbf{B}(\boldsymbol{\mu} - \mathbf{X}^\varepsilon(t))dt + \sqrt{\varepsilon}\,\boldsymbol{\Gamma} d\mathbf{W}(t) \tag{2.18}$$

i.e., the functions $\mathbf{b}(\cdot)$ and $\boldsymbol{\Gamma}(\cdot)$ in the SDE (2.8) are $\mathbf{b}(\mathbf{x}) = \mathbf{B}(\boldsymbol{\mu} - \mathbf{x})$ and $\boldsymbol{\Gamma}(\mathbf{x}) = \boldsymbol{\Gamma}$, where $\mathbf{B} = \mathrm{diag}(\{b_i\}), \boldsymbol{\Gamma} = \mathrm{diag}(\{\gamma_i\})$, and $b_i, \gamma_i > 0$ for all $i = 1, \ldots, m$. For this model, the capacity region can be expressed in closed form, as shown in the next Proposition.[6]

Proposition 2.1. *If $\mathbf{X}^\varepsilon(t)$ is defined by Eq. (2.18), then*

$$\tilde{\mathcal{R}}^{(cur)}_{\varepsilon,p} = \bigcap_{\ell \in \mathcal{E}'} \Big\{ \overline{\boldsymbol{\mu}} \in \mathbb{R}^N \,:\, |\nu_\ell| \leq 1 - \sqrt{\varepsilon \log(1/p) V_\ell \mathbf{M}_T V_\ell^\top} \Big\}.$$

In the particular case $\mathbf{B} = b\mathbf{I}$, Eq. (2.1) simplifies to

$$\tilde{\mathcal{R}}^{(cur)}_{\varepsilon,p} = \bigcap_{\ell \in \mathcal{E}'} \Big\{ \overline{\boldsymbol{\mu}} \in \mathbb{R}^N \,:\, |\nu_\ell| \leq 1 - \beta_\ell \Big\}. \tag{2.19}$$

Here, $\mathbf{M}_t = \boldsymbol{\Gamma}^2 \mathbf{B}^{-1}(\mathbf{I} - e^{-2\mathbf{B}t}) e^{\mathbf{B}(t-T)}$ and

$$\beta_\ell := \sqrt{\frac{(1 - e^{-2bT})\varepsilon \log(1/p)\sigma_\ell^2}{b}}, \quad \sigma_\ell^2 := V_\ell \boldsymbol{\Gamma}^2 V_\ell^\top.$$

We make the following remarks regarding Proposition 2.1: (i) $\tilde{\mathcal{R}}^{(\mathrm{cur})}_{\varepsilon,p}$ is a closed convex set; in particular, it is a polyhedron in $\mathbb{R}^N$. We note that this property enables us to incorporate the capacity region in OPF problems; (ii) β_ℓ is a strictly decreasing function of b, implying that $\tilde{\mathcal{R}}^{(\mathrm{cur})}_{\varepsilon,p}$ shrinks as b becomes smaller. This is intuitive, since for small values of b, the OU process will revert to its long-term mean $\boldsymbol{\mu}$ with less force; (iii) the longer the time T between two control instants, the greater the probability that the fluctuations in the power injections will result in an overload, yielding a smaller $\tilde{\mathcal{R}}^{(\mathrm{cur})}_{\varepsilon,p}$; (iv) the expression for $\tilde{\mathcal{R}}^{(\mathrm{cur})}_{\varepsilon,p}$ encloses in a single formula the dependency on the initial condition $\boldsymbol{\nu}$, on the window length T, and on the topology of the network, the physical properties of the transmission lines and the evolution of the stochastic power injections, encoded in the matrices $\mathbf{V}, \boldsymbol{\Gamma}, \mathbf{B}$.

[5] The Gaussianity assumption for wind power is debatable. While consistent with atmospheric physics [16] and recent wind park statistics [106, 12], different models are preferred for different timescales [121].

[6] Note that our framework allows to extend Proposition 2.1 to mixtures of OUs, providing flexibility to the modeler while keeping the benefits of closed-form expressions.

2.3 Capacity regions based on temperature overload

Since temperature responds gradually to current, a current overload of short duration does not necessarily imply an overload in temperature. By explicitly capturing the transient relationship between temperature and current, we can enlarge the conservative capacity region obtained in Section 2.2. In the following, we first provide a large deviation principle for the temperature overload event $\mathbb{P}(\|\boldsymbol{\Theta}^{\varepsilon,\boldsymbol{\tau}}\|_\infty \geq 1)$. After that, we define the temperature-overload based capacity region and prove a convexity result for it, analogous to the result in Section 2.2.

However, due to the non-local in time relationship between current and temperature, the decay rate for the temperature process is hard to compute explicitly. As a result, the capacity region cannot be expressed in closed form for even the simplest diffusion models. To address this issue, we develop two approximations of the capacity region: the first is an inner bound, while the second is based on a first-order Taylor expansion of the decay rate around $\boldsymbol{\tau} = \mathbf{0}$. These approximations have the following appealing properties, which make them amenable to application in OPF formulations. Firstly, both approximations are supersets of the current-based capacity region $\tilde{\mathcal{R}}^{(\mathrm{cur})}_{\varepsilon,p}$. Secondly, they have the same computational complexity as $\tilde{\mathcal{R}}^{(\mathrm{cur})}_{\varepsilon,p}$. Thirdly, for the special case where the stochastic power injections are modeled by an OU process, both regions can be expressed in closed form (Sections 2.3.4, 2.3.5). Finally, both approximations are convex over the deterministic power injections.

2.3.1 Derivation of the capacity region

Thanks to the relationship in Eq. (2.5), the contraction principle yields that $\boldsymbol{\Theta}_{\varepsilon,\boldsymbol{\tau}}$ satisfies a SPLDP with good rate function

$$\mathcal{I}_{\mathrm{tmp},\boldsymbol{\tau}}(\mathbf{h}) = \inf_{\substack{\mathbf{f}\in H^1_{\boldsymbol{\nu}}:\\ \xi_{\boldsymbol{\tau}}(\mathbf{f})=\mathbf{h}}} \mathcal{I}_{\mathrm{cur}}(\mathbf{f}) = \inf_{\substack{\mathbf{f}\in \mathbf{y}+\mathbf{V}H^1_{\boldsymbol{\mu}}:\\ \xi_{\boldsymbol{\tau}}(\mathbf{f})=\mathbf{h}}} \mathcal{I}_{\mathrm{cur}}(\mathbf{f}). \tag{2.20}$$

For the temperature overload event we thus have

$$\limsup_{\varepsilon\to 0} \varepsilon \log \mathbb{P}(\|\boldsymbol{\Theta}^{\varepsilon,\boldsymbol{\tau}}\|_\infty) \geq 1) \leq -\mathcal{I}^*_{\mathrm{tmp},\boldsymbol{\tau}}, \tag{2.21}$$

$$\mathcal{I}^*_{\mathrm{tmp},\boldsymbol{\tau}} = \inf_{\substack{\mathbf{h}\in\xi_{\boldsymbol{\tau}}(H^1_{\boldsymbol{\nu}})\\ \|\mathbf{h}\|_\infty\geq 1}} \mathcal{I}_{\mathrm{tmp},\boldsymbol{\tau}}(\mathbf{h}), \tag{2.22}$$

where $\mathcal{I}^*_{\mathrm{tmp},\boldsymbol{\tau}}$ is the *temperature decay rate*. Letting, for $\ell \in \mathcal{E}'$,

$$\omega_\ell = \inf_{\substack{\mathbf{h}\in\xi_{\boldsymbol{\tau}}(H^1_{\boldsymbol{\nu}}):\\ \|h_\ell\|_\infty\geq 1}} \mathcal{I}_{\mathrm{tmp},\boldsymbol{\tau}}(\mathbf{h}) = \inf_{\substack{\mathbf{g}\in H^1_{\boldsymbol{\mu}}:\\ \|\xi_{\tau_\ell}(y_\ell+V_\ell\mathbf{g})\|_\infty\geq 1}} \mathcal{I}_{\mathrm{pow}}(\mathbf{g}), \tag{2.23}$$

we see that the decay rate for the temperature is $\mathcal{I}^*_{\text{tmp},\boldsymbol{\tau}} = \min_{\ell\in\mathcal{E}'} \omega_\ell$. Note that ω_ℓ and $\mathcal{I}^*_{\text{tmp},\boldsymbol{\tau}}$ depend on $\overline{\boldsymbol{\mu}}, \boldsymbol{\tau}$ and T. As before, Eq. (2.21) yields the following approximation for the rare event probability, for small ε:

$$\mathbb{P}(\|\boldsymbol{\Theta}^{\varepsilon,\boldsymbol{\tau}}\|_\infty \geq 1) \approx e^{-\mathcal{I}^*_{\text{tmp},\boldsymbol{\tau}}(\overline{\boldsymbol{\mu}})/\varepsilon}. \tag{2.24}$$

This leads to the following definition of the capacity region

$$\begin{aligned}\tilde{\mathcal{R}}^{(\text{tmp},\boldsymbol{\tau})}_{\varepsilon,p} &:= \{\overline{\boldsymbol{\mu}} \in \mathbb{R}^N \,:\, \mathcal{I}^*_{\text{tmp},\boldsymbol{\tau}}(\overline{\boldsymbol{\mu}}) \geq -\varepsilon \log(p)\} \\ &= \bigcap_{\ell\in\mathcal{E}'} \{\overline{\boldsymbol{\mu}} \in \mathbb{R}^N \,:\, \omega_\ell(\overline{\boldsymbol{\mu}}) \geq -\varepsilon \log(p)\}.\end{aligned} \tag{2.25}$$

We have the following convexity result:

Lemma 2.4. *$\tilde{\mathcal{R}}^{(tmp,\boldsymbol{\tau})}_{\varepsilon,p}$ is convex in the deterministic power injections vector $\boldsymbol{\mu}_D$.*

The variational problem for the temperature overload (2.22) is difficult to solve in general, and numerics can also prove to be challenging. Motivated by this difficulty, in the next subsection we develop approximations for the temperature decay rate, and the corresponding capacity regions, by reducing the problem (2.22) to the easier problem (2.15).

2.3.2 Inner bound for the capacity region

In this section, we develop an inner bound for the capacity region $\tilde{\mathcal{R}}^{(\text{tmp},\boldsymbol{\tau})}_{\varepsilon,p}$ which is larger than the capacity region $\tilde{\mathcal{R}}^{(\text{cur})}_{\varepsilon,p}$ based on current overload, and thus captures some of the benefit of incorporating temperature dynamics. Define

$$\mathcal{I}^{(LB)}_{\text{tmp},\boldsymbol{\tau}} := \min_{\ell\in\mathcal{E}'} \psi^{(\alpha_\ell)}_\ell \wedge \psi^{(-\alpha_\ell)}_\ell.$$

The next lemma shows that $\mathcal{I}^{(LB)}_{\text{tmp},\boldsymbol{\tau}}$ is a lower bound for the temperature decay rate, i.e., $\mathcal{I}^*_{\text{tmp},\boldsymbol{\tau}} \geq \mathcal{I}^{(LB)}_{\text{tmp},\boldsymbol{\tau}}$.

Lemma 2.5. *For all $\ell \in \mathcal{E}'$, we have $\omega_\ell \geq \psi^{(\alpha_\ell)}_\ell \wedge \psi^{(-\alpha_\ell)}_\ell$, where $\alpha_\ell = \sqrt{\frac{1-\nu_\ell^2 e^{-T/\tau_\ell}}{1-e^{-T/\tau_\ell}}}$.*

The capacity region based on the lower bound $\mathcal{I}^{(LB)}_{\text{tmp},\boldsymbol{\tau}}$ is

$$\tilde{\mathcal{R}}^{(\text{tmp},\boldsymbol{\tau},LB)}_{\varepsilon,p} := \{\overline{\boldsymbol{\mu}} \in \mathbb{R}^N \,:\, \mathcal{I}^{(LB)}_{\text{tmp},\boldsymbol{\tau}}(\boldsymbol{\tau},\overline{\boldsymbol{\mu}}) \geq -\varepsilon \log(p)\}.$$

The following proposition states that the capacity region based on the lower bound, while being an inner approximation of the actual temperature-based region, is less conservative than the current-based capacity constraint.

Proposition 2.2. *$\mathcal{I}^*_{cur} \leq \mathcal{I}^{(LB)}_{tmp,\boldsymbol{\tau}} \leq \mathcal{I}^*_{tmp,\boldsymbol{\tau}}$ and $\tilde{\mathcal{R}}^{(cur)}_{\varepsilon,p} \subseteq \tilde{\mathcal{R}}^{(tmp,\boldsymbol{\tau},LB)}_{\varepsilon,p} \subseteq \tilde{\mathcal{R}}^{(tmp,\boldsymbol{\tau})}_{\varepsilon,p}$ for all $\boldsymbol{\tau} \geq \mathbf{0}$.*

As a consequence, using $\tilde{\mathcal{R}}^{(\text{tmp},\boldsymbol{\tau},LB)}_{\varepsilon,p}$ over $\tilde{\mathcal{R}}^{(\text{cur})}_{\varepsilon,p}$ allows for larger power injections values (i.e., less curtailment), while still bounding the probability of a temperature overload and without additional computational burden. Finally, we note that the inner bound satisfies the following convexity property.

Lemma 2.6. *$\tilde{\mathcal{R}}^{(tmp,\boldsymbol{\tau},LB)}_{\varepsilon,p}$ is convex in the deterministic power injections vector $\boldsymbol{\mu}_D$.*

The proof goes along the same lines of the proofs of Lemmas 2.2 and 2.4 and is therefore omitted.

2.3.3 Taylor approximation of the decay rate and corresponding capacity region

In this section we derive a heuristic approximation for the temperature decay rate

$$\mathcal{I}^*_{\text{tmp},\boldsymbol{\tau}} = \inf_{\mathbf{h}\in\xi_{\boldsymbol{\tau}}(\mathbf{y}+\mathbf{V}H^1_{\boldsymbol{\mu}}),\, \|\mathbf{h}\|_\infty \geq 1} \mathcal{I}_{\text{tmp},\boldsymbol{\tau}}(\mathbf{h}), \tag{2.26}$$

based on a Taylor expansion around $\boldsymbol{\tau} = \mathbf{0}$. First, write the temperature rate function in Eq. (2.20) as

$$\mathcal{I}_{\text{tmp},\boldsymbol{\tau}}(\mathbf{h}) = \begin{cases} G(\boldsymbol{\tau},\mathbf{h}) & \text{if } \mathbf{h} \in \xi_{\boldsymbol{\tau}}(\mathbf{y}+\mathbf{V}H^1_{\boldsymbol{\mu}}), \\ \infty & \text{otherwise,} \end{cases}$$

where $G(\boldsymbol{\tau},\mathbf{h})$ is defined explicitly in Eq. (2.46) in Appendix 2.A.

Taylor approximation 1. *Let $\mathbf{f}_*$ be the optimal current path to overflow. For small $\boldsymbol{\tau}$, we will use the approximation*

$$\mathcal{I}^*_{tmp,\boldsymbol{\tau}} \approx \mathcal{I}^{(TL)}_{tmp,\boldsymbol{\tau}} := \mathcal{I}^*_{cur} + \boldsymbol{\tau}\cdot\nabla_{\boldsymbol{\tau}} G(\boldsymbol{\tau},\mathbf{f}^2_*)|_{\boldsymbol{\tau}=\mathbf{0}}, \tag{2.27}$$

where G is defined in Eq. (2.46) in Appendix 2.A.
If $\boldsymbol{\tau}$ is of the form $\boldsymbol{\tau} = \tau_0(1,\ldots,1)^\top$, $\tau_0 > 0$, we obtain the closed-form expression

$$\mathcal{I}^{(TL)}_{tmp,\boldsymbol{\tau}} = \mathcal{I}^*_{cur} + \tau_0 \Phi_{\mathbf{f}_*}, \tag{2.28}$$

where

$$\Phi_{\mathbf{f}_*} := \sum_{i=1}^{n_w} \Big[K_i(\mathbf{f}_*(T), \mathbf{f}'_*(T)) - K_i(\mathbf{f}_*(0), \mathbf{f}'_*(0)) \Big], \tag{2.29}$$

$$K_i(\mathbf{f}_*(t), \mathbf{f}'_*(t)) := \frac{1}{2}\Big(\frac{V_i^+ \mathbf{f}'_*(t) - b_i(V_i^+(\mathbf{f}_*(t)-\mathbf{y}))}{\gamma_i(V_i^+(\mathbf{f}_*(t)-\mathbf{y}))} \Big)^2, \tag{2.30}$$

with $b_i(\cdot)$ and $\gamma_i(\cdot)$ the drift and diffusion terms for the i-th stochastic power injection X_i^ε in Eq. (2.8)*. In particular the approximation $\mathcal{I}^{(TL)}_{tmp,\boldsymbol{\tau}}$ depends only on the current decay rate $\mathcal{I}^*_{cur}$ and on the values $\mathbf{f}_*(0), (\mathbf{f}_*)'(0), \mathbf{f}_*(T), (\mathbf{f}_*)'(T)$.*

The heuristic is motivated by the formal Taylor expansion of $I^*_{t,\boldsymbol{\tau}}$ around $\tau = 0$, i.e., $\mathcal{I}^*_{t,0} + \boldsymbol{\tau} \cdot \nabla_{\boldsymbol{\tau}} I^*_{t,\boldsymbol{\tau}}|_{\boldsymbol{\tau}=\mathbf{0}} + o(\boldsymbol{\tau})$. If $\boldsymbol{\tau} = \mathbf{0}$, the optimal temperature path to overflow is $\mathbf{h}_* = (\mathbf{f}_*)^2$, so $\mathcal{I}^*_{\text{tmp},\boldsymbol{\tau}} = \mathcal{I}_{\text{cur}^2}(\mathbf{h}_*) = \mathcal{I}_{\text{cur}}(\mathbf{f}_*) = \mathcal{I}^*_{\text{cur}}$, and the substitution of $\nabla_{\boldsymbol{\tau}} G(\boldsymbol{\tau}, \mathbf{f}_*^2)|_{\boldsymbol{\tau}=\mathbf{0}}$ for $\nabla_{\boldsymbol{\tau}} \mathcal{I}^*_{\text{tmp},\boldsymbol{\tau}}|_{\boldsymbol{\tau}=\mathbf{0}}$ is motivated by an infinite-dimensional version of Danskin's Theorem [22], Proposition 4.13. The explicit calculations for the case $\boldsymbol{\tau} = \tau_0(1, \dots, 1)^\top$, $\tau_0 > 0$, are reported in Appendix 2.A.

Eq. (2.28) provides an approximation of the temperature decay rate which depends only on the current decay rate and the corresponding optimal path, which are generally easier to obtain. The capacity region corresponding to the Taylor approximation is

$$\tilde{\mathcal{R}}^{(\text{tmp},\boldsymbol{\tau},TL)}_{\varepsilon,p} := \{\overline{\boldsymbol{\mu}} \in \mathbb{R}^N \,:\, \mathcal{I}^*_{\text{cur}}(\bar{\boldsymbol{\mu}}) + \tau_0 \Phi_{\mathbf{f}_*} \geq -\varepsilon \log(p)\}.$$

In Section 2.3.5 we show that the inequalities $\mathcal{I}^{(TL)}_{\text{tmp},\boldsymbol{\tau}} \geq \mathcal{I}^*_{\text{cur}}$ and $\tilde{\mathcal{R}}^{(t,\boldsymbol{\tau},TL)}_{\varepsilon,p} \supseteq \tilde{\mathcal{R}}^{(\text{cur})}_{\varepsilon,p}$ hold in the OU case, confirming the intuition that the temperature-based approach is less conservative than the current-based one.

2.3.4 Explicit computations for OU: lower bound

In this section we assume that the power injection process $\mathbf{X}^\varepsilon(t)$ follows the OU process in Eq. (2.18), and we explicitly compute the lower bound $\mathcal{I}^{(LB)}_{\text{tmp},\boldsymbol{\tau}}$ and the corresponding capacity region $\tilde{\mathcal{R}}^{(\text{tmp},\boldsymbol{\tau},LB)}_{\varepsilon,p}$.

Proposition 2.3. *If $\mathbf{X}^\varepsilon(t)$ is defined by Eq.* (2.18)*, then*

$$\mathcal{I}^{(LB)}_{tmp,\boldsymbol{\tau}} = \min_{\ell \in \mathcal{E}'} \frac{(\alpha_\ell - |\nu_\ell|)^2}{V_\ell \mathbf{M}_T V_\ell^\top},$$

$$\tilde{\mathcal{R}}^{(tmp,\boldsymbol{\tau},LB)}_{\varepsilon,p} = \bigcap_{\ell \in \mathcal{E}'} \{\overline{\boldsymbol{\mu}} \in \mathbb{R}^N : \frac{(\alpha_\ell - |\nu_\ell|)^2}{V_\ell \mathbf{M}_T V_\ell^\top} \geq -\varepsilon \log(p)\},$$

where

$$\alpha_\ell = \sqrt{\frac{1 - \nu_\ell^2 e^{-T/\tau_\ell}}{1 - e^{-T/\tau_\ell}}}, \quad \mathbf{M}_T = \boldsymbol{\Gamma}^2 \mathbf{B}^{-1} (\mathbf{I} - e^{-2\mathbf{B}t}) e^{\mathbf{B}(t-T)}.$$

In the particular case $\mathbf{B} = b\mathbf{I}$, we have

$$\tilde{\mathcal{R}}^{(tmp,\boldsymbol{\tau},LB)}_{\varepsilon,p} := \bigcap_{\ell \in \mathcal{E}'} \{\overline{\boldsymbol{\mu}} \in \mathbb{R}^N \,:\, |\nu_\ell| \leq \delta_\ell\}, \tag{2.31}$$

$$\delta_\ell = \sqrt{1 - \eta_\ell^2 e^{-T/\tau_\ell}(1 - e^{-T/\tau_\ell})} - \eta_\ell(1 - e^{-T/\tau_\ell}), \tag{2.32}$$

$$\eta_\ell := \sqrt{\frac{\varepsilon \log(1/p)\sigma_\ell^2(1 - e^{-2bT})}{b}} < 1,\ \sigma_\ell^2 = V_\ell \mathbf{\Gamma}^2 V_\ell^\top. \tag{2.33}$$

If $\mathbf{B} = b\mathbf{I}$, we see from Prop. 2.3 that $\tilde{\mathcal{R}}_{\varepsilon,p}^{(\text{tmp},\boldsymbol{\tau},LB)}$ is a convex polyhedron in $\mathbb{R}^N$, as in the case of the current region, and is in particular a scaled version of the polyhedron $\tilde{\mathcal{R}}_{\varepsilon,p}^{(\text{cur})}$. Moreover, $\delta_\ell \in (1 - \eta_\ell, 1), \delta_\ell \xrightarrow{\boldsymbol{\tau}\to\infty} 1$ and $\delta_\ell \xrightarrow{\boldsymbol{\tau}\to 0} \eta_\ell$. This means that, as $\boldsymbol{\tau}$ increases, the capacity region (2.31) gets closer to the larger region $\{\overline{\boldsymbol{\mu}} \in \mathbb{R}^N : \|\boldsymbol{\nu}\|_\infty < 1\}$, which is the stability region for a deterministic system. On the other hand, as $\boldsymbol{\tau} \to 0$, the region in Eq. (2.31) boils down to the smaller current-based capacity region given in Eq. (2.1).

2.3.5 Explicit computations for OU: Taylor approximation

In this section we consider again the OU process $\mathbf{X}^{(\varepsilon)}$ in Eq. (2.18) in the particular case $\mathbf{B} = b\mathbf{I}$, and we develop the capacity regions based on the Taylor approximation 1.

Proposition 2.4. *For $\boldsymbol{\tau} = \tau_0(1, \dots, 1)^\top$ we have*

$$\mathcal{I}_{tmp,\boldsymbol{\tau}}^{(TL)} = (1 + 2\tau_0 b)\,\mathcal{I}_{cur}^*(\bar{\boldsymbol{\mu}}) = (1 + 2\tau_0 b) \min_{\ell\in\mathcal{E}'} \frac{(1 - |\nu_\ell|)^2}{V_\ell \mathbf{M}_T V_\ell^\top}, \tag{2.34}$$

$$\tilde{\mathcal{R}}_{\varepsilon,p}^{(tmp,\boldsymbol{\tau},TL)} = \bigcap_{\ell\in\mathcal{E}'} \Big\{\overline{\boldsymbol{\mu}} \in \mathbb{R}^N : |\nu_\ell| \le 1 - \eta_\ell/\sqrt{1 + 2\tau_0 b}\Big\},$$

It is clear that $\tilde{\mathcal{R}}_{\varepsilon,p}^{(\text{tmp},\boldsymbol{\tau},TL)}$ is a convex polyhedron, as it was the case for the current region $\tilde{\mathcal{R}}_{\varepsilon,p}^{(\text{cur})}$ and the lower bound region $\tilde{\mathcal{R}}_{\varepsilon,p}^{(t,\tau_0,LB)}$. Moreover, since $1 + 2\tau_0 b > 0$, we see that $\tilde{\mathcal{R}}_{\varepsilon,p}^{(\text{tmp},\boldsymbol{\tau},TL)} \supseteq \tilde{\mathcal{R}}_{\varepsilon,p}^{(\text{cur})}$ and in particular $\tilde{\mathcal{R}}_{\varepsilon,p}^{(\text{tmp},\boldsymbol{\tau},TL)}$ is a scaled version of $\tilde{\mathcal{R}}_{\varepsilon,p}^{(\text{cur})}$. Recall that this was also the case for the lower bound capacity region: the difference is that, while the lower bound holds for every $\boldsymbol{\tau} > 0$, the approximation $\mathcal{I}_{\text{tmp},\boldsymbol{\tau}}^{(TL)}$ is meaningful only for small τ_0. In general, $\tilde{\mathcal{R}}_{\varepsilon,p}^{(\text{tmp},\boldsymbol{\tau},TL)}$ and $\tilde{\mathcal{R}}_{\varepsilon,p}^{(\text{tmp},\boldsymbol{\tau},LB)}$ are not subsets of each other.

2.4 Numerics

In order to compute the temperature decay rate $\mathcal{I}_{\text{tmp},\boldsymbol{\tau}}^*$, one has to solve the variational problem in Eq. (2.23), which is computationally harder than the one for the current in Eq. (2.15), due to the integral mapping in Eq. (2.6).

On the other hand, the theory we presented enables us to reduce the computation of the decay rates $\mathcal{I}^{(LB)}_{\text{tmp},\boldsymbol{\tau}}, \mathcal{I}^{(TL)}_{\text{tmp},\boldsymbol{\tau}}$ to the easier variational problem for $\mathcal{I}^{*}_{\text{cur}}$, capturing the benefits of incorporating the temporal dynamics between current and temperature without additional cost. Variational problems like Eq. (2.15), which are based on Freidlin-Wentzell theory, are well studied in the literature, and when closed-form expression are not available efficient numerical algorithms have been developed [82].

In the next subsections we apply our theory to derive the capacity regions for two IEEE test cases in MATPOWER [210], and we quantify the capacity gains achieved by $\tilde{\mathcal{R}}^{(\text{tmp},\boldsymbol{\tau},LB)}_{\varepsilon,p}, \tilde{\mathcal{R}}^{(\text{tmp},\boldsymbol{\tau},TL)}_{\varepsilon,p}$ over $\tilde{\mathcal{R}}^{(\text{cur})}_{\varepsilon,p}$ assuming an OU model for power injections.

Thanks to the analytic characterization of capacity regions for the OU model, our approach is fully scalable and can effortlessly be applied to larger power networks, as there is virtually no computational burden in computing $\tilde{\mathcal{R}}^{(\text{cur})}_{\varepsilon,p}$ and, therefore, all the other capacity regions. For a detailed analysis on computational costs for solving Eq. (2.15) for a general SDE, the interested reader is referred to [82], Section 3.3.

2.4.1 IEEE 14-bus test network

In this section we develop capacity regions for the IEEE 14-bus test network, corresponding to the test case *case14* in [210]. The grid consists of 14 nodes and 20 lines, and the original test case has constant deterministic power injections $P_D \in \mathbb{R}^{14}$. We replace two of the deterministic injections (nodes 2 and 13) by OU processes with long term mean equal to the original deterministic power injection, and we assume we control the injections at nodes 6 and 9. The test case reports the parameters P_D, x_{ij} and $\widetilde{C}$, but does not include line limits, which we define as follows. For each line ℓ we set the maximum permissible current $I_{\ell,\max} = K|I_{\ell,\text{nom}}|$, $I_{\ell,\text{nom}}$ being the nominal current in line ℓ obtained from P_D via the DC power flow equation, and $K = 1.5$. We set $T = 1, b_i = 1, \gamma_i^2 = 10, \varepsilon = 0.25$ and $\tau_0 = 0.5$.

We compute two-dimensional capacity regions, which correspond to the amount of power that can be injected at the controllable sources so that the probability of overload in $[0, T]$ is sufficiently small. The current-based capacity region is

$$\begin{aligned}\tilde{\mathcal{R}}^{(\text{cur})}_{\varepsilon,p} = \{(\overline{\mu}_6, \overline{\mu}_9) \in \mathbb{R}^2 \,|,\, \overline{\boldsymbol{\mu}} = (P_{D,2}, \ldots, P_{D,5}, \overline{\mu}_6, P_{D,7},\\ P_{D,8}, \overline{\mu}_9, P_{D,10}, \ldots, P_{D,14}),\, \mathcal{I}_{\text{cur}}(\overline{\boldsymbol{\mu}}) \geq -\epsilon \log(p)\},\end{aligned}$$

and the other regions are defined similarly. In Figs. 2.1a, 2.1b the 2-dimensional capacity regions $\tilde{\mathcal{R}}^{(\text{cur})}_{\varepsilon,p}, \tilde{\mathcal{R}}^{(\text{tmp},\boldsymbol{\tau},LB)}_{\varepsilon,p}$ and $\tilde{\mathcal{R}}^{(\text{tmp},\boldsymbol{\tau},TL)}_{\varepsilon,p}$ (denoted as $\tilde{\mathcal{R}}^{(\text{cur})}, \tilde{\mathcal{R}}^{(\text{LB})}$

and $\tilde{\mathcal{R}}^{(\mathrm{TL})}$ in the legend) are shown for two different target probabilities p, together with the region corresponding to a deterministic system

$$\mathcal{R}_{\mathrm{det}} = \{(\mu_1, \mu_2) \in \mathbb{R}^2 \,|\, |\nu_\ell| \leq 1 \quad \forall \ell = 1, 2, 3\} \tag{2.35}$$

In particular, Fig. 2.1b shows that for $p = 10^{-7}$ the lower bound region $\tilde{\mathcal{R}}_{\varepsilon,p}^{(\mathrm{tmp},\boldsymbol{\tau},LB)}$ is more than two times bigger than $\tilde{\mathcal{R}}_{\varepsilon,p}^{(\mathrm{cur})}$, and the Taylor region $\tilde{\mathcal{R}}_{\varepsilon,p}^{(\mathrm{tmp},\boldsymbol{\tau},TL)}$ is approximately two times bigger than $\tilde{\mathcal{R}}_{\varepsilon,p}^{(\mathrm{tmp},\boldsymbol{\tau},LB)}$. This result suggests that for small target probabilities, the temperature-based approach yields a significative capacity gain.

Another application of the proposed methodology is the identification of the most vulnerable parts of the grid. For a given value of $\overline{\boldsymbol{\mu}}$, let $\ell^*(\overline{\boldsymbol{\mu}}) := \mathrm{argmin}_{\ell \in \mathcal{E}'} \psi_\ell^{(1)}(\overline{\boldsymbol{\mu}}) \wedge \psi_\ell^{(-1)}(\overline{\boldsymbol{\mu}})$ denote the line with the highest chance of overloading in Eq. (2.16), and, for a line $k \in \mathcal{E}'$, define

$$S_k := \{(\overline{\mu}_6, \overline{\mu}_9) \in \mathbb{R}^2 \,|\, \|\boldsymbol{\nu}\|_\infty \leq 1, \ell^*(\overline{\boldsymbol{\mu}}) = k\} \subset \mathcal{R}_{\mathrm{det}}. \tag{2.36}$$

The region $S_k \subset \mathcal{R}_{\mathrm{det}}$ characterizes the *controllable* power injections such that, in the event of large fluctuations of stochastic power injections, line k is the most likely line to overload. The S_k-s partition $\mathcal{R}_{\mathrm{det}}$ in several subregions, as shown in Fig. 2.1c. Such characterization can help detecting the most vulnerable components of the grid: in this case, line $(12, 13)$, corresponding to the biggest sub-region in Fig. 2.1c. Finally, Fig. 2.1d shows the topology of the network.

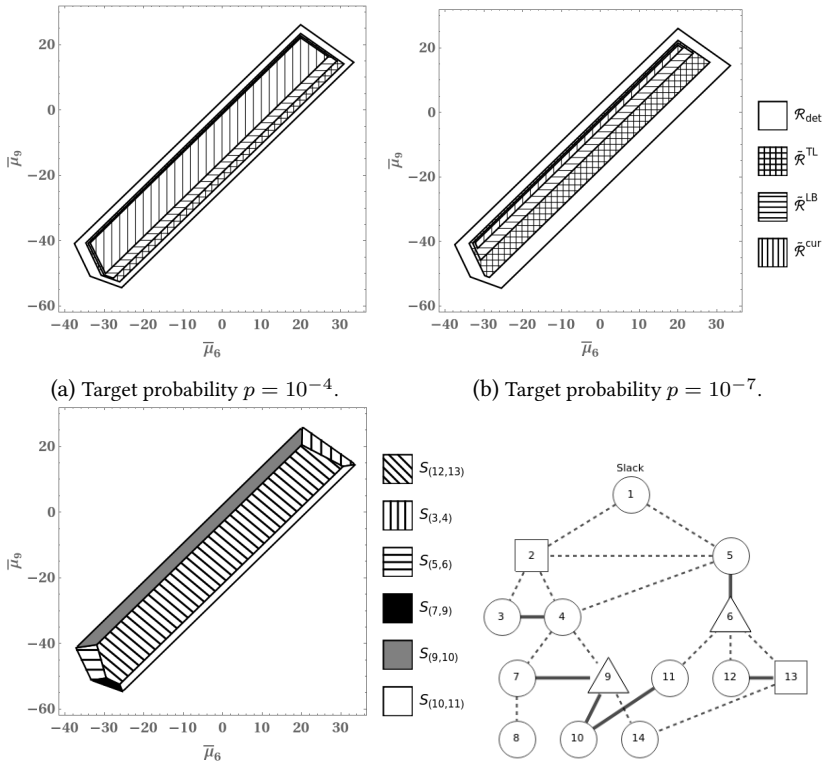
2.4.2 IEEE 118-bus test network

In this section we perform a case study on a larger system, corresponding to the test case *c118swf.m* [130]. The system has 118 nodes, 210 lines and 52 generators, 11 of which are modeled as wind units (indexed by $j_1, \ldots, j_{11}$). In order to simulate a more heavily loaded system, we define $I_{\max}$ to be equal to 50% of the line limits provided in the test case.

For our study, we first solve a DC OPF [177], which is an optimization problem determining the generation schedule that minimizes the total system generation cost, while satisfying demand/supply balance and network physical constraints, under the assumptions of the DC approximations (see Section 1.3.1). Let $\bar{\boldsymbol{\mu}} \in \mathbb{R}^{118}$ be the resulting optimal net power injections vector.

Next, we model the 11 wind generators as OU processes, using the hour as the unit for temporal quantities. The parameter μ_k of generator j_k is set to be equal to $\bar{\mu}_{j_k}$, which is interpreted as the *nominal power injection* of generator j_k.

The parameters ε, $\mathbf{B} = b\mathbf{I}, \boldsymbol{\Gamma} = \mathrm{diag}(\{\gamma_I\})$ and T are calibrated in such a way that the standard deviation of each OU process at the end time T matches



(a) Target probability $p = 10^{-4}$.

(b) Target probability $p = 10^{-7}$.

(c) Partition of $\mathcal{R}_{\text{det}}$.

(d) Network topology.

Figure 2.1: a,b) Capacity regions for the IEEE 14-bus network, depicted using different mesh styles, for two different target probabilities; c) Subdivision of $\mathcal{R}_{\text{det}}$ according to which lines are the most vulnerable, as in Eq. (2.36); d) IEEE-14 topology. Stochastic and (deterministic) controllable nodes are represented with square and triangular vertexes, respectively. The six solid lines are the most vulnerable ones.

realistic values for wind power forecasting error (expressed as a fraction of the wind plant installed capacity) over different control periods:

$$\text{std}_{j_k}(T) = \sqrt{\frac{\varepsilon l_k^2}{2b}(1 - e^{-2Tb})} = q(T) \cdot \mu_{j_k}^{(\text{installed})}. \tag{2.37}$$

Given T, we set $q = q(T), b = 1, \varepsilon = 1$ and solve Eq. (2.37) for γ_k. The values for $q(T)$, shown in Fig. 2.2a,2.2b, are taken from [89], and correspond to the Root Mean Squared Forecast Error obtained applying a persistence forecast to ERCOT wind data. Note that this setting can capture renewable generators with

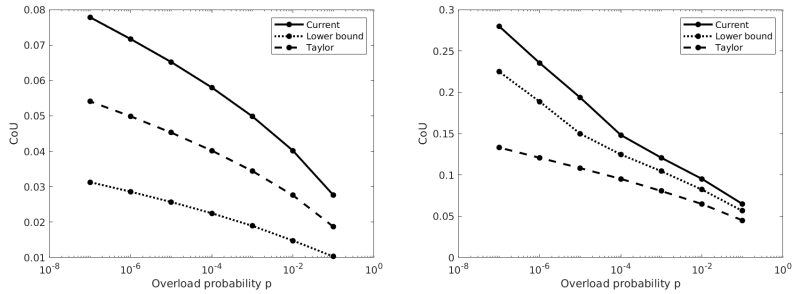
different installed capacities. The overload probabilities are chosen in the range $[10^{-7}, 10^{-1}]$, and $\tau = 0.5$.

To quantify the capacity gain achieved by the different regions, for each choice of the parameters we solve three distinct DC OPFs, each incorporating a different capacity region $\mathcal{R}$ in the constraints. Note that since the capacity regions are convex polytopes, solving these OPFs has the same computational cost as solving the deterministic one.

Next, we compare the total system costs, which is the value of the objective function at optimality, to the cost obtained by solving the deterministic OPF (that is, the one incorporating $\mathcal{R}_{\text{det}}$ in the constraints), by means of the *Cost of Uncertainty* (CoU) metric

$$\text{CoU}^{(\mathcal{R})}(q,p) = \frac{\text{cost}^{\mathcal{R}}(q,p) - \text{cost}^{\text{det}}}{\text{cost}^{\text{det}}} \geq 0, \tag{2.38}$$

defined as the relative increase in system costs when uncertainty-aware reliability constraints are considered. Fig. 2.2 reports $\text{CoU}(q,p)$ for various values of q and p.



(a) $T = 1/4$ (15 minutes), $q(T) = 0.018$. (b) $T = 1$ (60 minutes), $q(T) = 0.04$.

Figure 2.2: $\text{CoU}^{(\mathcal{R})}$ for different overload probabilities p and time intervals T, $\mathcal{R} \in \{\tilde{\mathcal{R}}_{\varepsilon,p}^{(\text{cur})}, \tilde{\mathcal{R}}_{\varepsilon,p}^{(\text{tmp},\boldsymbol{\tau},LB)}, \tilde{\mathcal{R}}_{\varepsilon,p}^{(\text{tmp},\boldsymbol{\tau},TL)}\}$.

We see that enforcing constraints on line currents results in higher system costs than the ones achieved by using temperature-based constraints, consistently across different probability levels and time intervals. The gain is more pronounced over shorter intervals, capturing the intuition that current overloads are permissible for short periods, and for smaller probabilities: for instance, CoU drops from 8% to 3% when $\tilde{\mathcal{R}}_{\varepsilon,p}^{(\text{tmp},\boldsymbol{\tau},LB)}$ is used over $\tilde{\mathcal{R}}_{\varepsilon,p}^{(\text{cur})}$, for $T = 1/4, p = 10^{-7}$.

2.5 Concluding remarks

We employed large deviations theory to develop tractable capacity regions for power grids with variable power injections, modeled as small-noise diffusion processes, assuming currents behave according to the DC power flow equations. These capacity regions define the set of initial power injections such that the probability of a current/temperature overload in a given interval is very small, and can be used as computationally tractable chance-constraints in OPF formulations. Incorporating the transient relationship between line temperature and line current leads to enlarged capacity regions, due to the fact that a temporary current overload does not necessarily lead to a temperature overload. While this enlarged region is difficult to compute, we develop tractable approximations that improve upon the capacity region defined by the conservative current overload constraint. Moreover, we note that potential of our large-deviations approach goes beyond the development of capacity regions. For example, our results can be used to speed up more detailed simulations, as in [192].

The results presented in this chapter are valid in the asymptotic small-noise regime $\varepsilon \to 0$. Analytic chance constraints that are valid in the prelimit as well are developed in Chapter 3. Finally, the ranking of transmission lines according to their overload probability makes our techniques applicable to identify the most vulnerable parts of the network (as in Fig. 2.1d). This topic is further pursued in Chapter 4, with applications to a large-scale realistic network.

Appendix

2.A Extended proofs

Proof of Lemma 2.1. Recall that $\overline{\mathbf{V}} = \mathbf{\Lambda}\mathbf{D}\mathbf{A}\overline{\mathbf{L}}$, where

$$\mathbf{\Lambda} = \mathrm{diag}(1/I_{1,\max}, \ldots, 1/I_{m,\max}),$$

and $\mathbf{D}, \mathbf{A}$ and $\overline{\mathbf{L}}$ have been defined in Section 1.2.2. Since $x_\ell \neq 0$, $I_{\ell,\max} \neq 0$ for all $\ell \in \mathcal{E}$, the matrices $\mathbf{\Lambda}$ and $\mathbf{D}$ are nonsingular. Following [15] we see that rank $\overline{\mathbf{L}} = N$, and Lemma 2.2 in [9] guarantees that rank $(\mathbf{A}) = N$ and Ker $(\mathbf{A}) = \mathrm{Span}((1, \ldots, 1)^T)$. Since $\mathbf{D}$ is nonsingular,

$$\mathrm{rank}\,(\mathbf{D}\mathbf{A}\overline{\mathbf{L}}) = \mathrm{rank}\,(\mathbf{A}\overline{\mathbf{L}}) \leq \min(\mathrm{rank}\,(\mathbf{A}), \mathrm{rank}\,(\overline{\mathbf{L}})) = N.$$

On the other hand, if $\mathbf{x} \in$ Ker $(\mathbf{A}\overline{\mathbf{L}})$ then $\mathbf{A}\overline{\mathbf{L}}\mathbf{x} = 0 \iff \overline{\mathbf{L}}\mathbf{x} \in$ Ker $(\mathbf{A}) =$ $\mathrm{Span}((1, \ldots, 1)^T) \iff \overline{\mathbf{L}}\mathbf{x} = 0 \iff \mathbf{x} \in$ Ker $(\overline{\mathbf{L}})$, where in the second implication we used that the first component of $\overline{\mathbf{L}}\mathbf{x}$ is 0. Therefore

dim Ker $(\mathbf{A}\overline{\mathbf{L}}) = \text{dim Ker}\,(\overline{\mathbf{L}}) = 1$, yielding

$$\text{rank}\,(\mathbf{A}\overline{\mathbf{L}}) = N + 1 - \text{dim Ker}\,(\mathbf{A}\overline{\mathbf{L}}) = N.$$

Since $\mathbf{\Lambda}$ is nonsingular, the matrix $\overline{\mathbf{V}}$ has N linear independent columns. But its first column is zero (since the first column of $\overline{\mathbf{L}}$ is zero), therefore the columns from 2 to $n_w \leq N$ of $\overline{\mathbf{V}}$ are linearly independent, hence $\mathbf{V}$ has full rank n_w. □

Proof of Lemma 2.2. First notice that a vector $(\boldsymbol{\mu}, \boldsymbol{\mu}_D)$ such that

$$\|\boldsymbol{\nu}\|_\infty = \|\mathbf{V}\boldsymbol{\mu} + \mathbf{V_D}\boldsymbol{\mu}_D\|_\infty < 1$$

belongs to the capacity region $\tilde{\mathcal{R}}^{(\text{cur})}_{\varepsilon,p}$, i.e.,

$$\mathcal{I}^*_{\text{cur}}(\boldsymbol{\mu}, \boldsymbol{\mu}_D) = \inf_{\substack{\mathbf{g}\in H^1_{\boldsymbol{\mu}}: \\ \|\mathbf{V}\mathbf{g}+\mathbf{V_D}\boldsymbol{\mu}_D\|_\infty = 1}} \mathcal{I}_{\text{pow}}(\mathbf{g}) \geq \varepsilon \log(1/p),$$

if and only if the following implication holds:

$$\exists \mathbf{g} \in H^1_{\boldsymbol{\mu}} \text{ s.t. } \mathcal{I}_{\text{pow}}(\mathbf{g}) < \varepsilon \log(1/p) \implies \|\mathbf{V} + \mathbf{V_D}\boldsymbol{\mu}_D\mathbf{g}\|_\infty < 1. \tag{2.39}$$

Consider two admissible vectors $(\boldsymbol{\mu}, \boldsymbol{\mu}_D), (\boldsymbol{\mu}, \tilde{\boldsymbol{\mu}}_\mathbf{D}) \in \tilde{\mathcal{R}}^{(\text{cur})}_{\varepsilon,p}$, and let $\lambda \in [0,1]$. We want to show that $(\boldsymbol{\mu}, \lambda\boldsymbol{\mu}_D + (1-\lambda))\tilde{\boldsymbol{\mu}}_\mathbf{D} \in \tilde{\mathcal{R}}^{(\text{cur})}_{\varepsilon,p}$. To this end, take $\mathbf{g} \in H^1_{\boldsymbol{\mu}}$ to be such that $\mathcal{I}_{\text{pow}}(\mathbf{g}) < \varepsilon \log(1/p)$, and let us write

$$\begin{aligned}
&\|\lambda\mathbf{V_D}\boldsymbol{\mu}_D + (1-\lambda)\mathbf{V_D}\tilde{\boldsymbol{\mu}}_\mathbf{D} + \mathbf{V}\mathbf{g}\|_\infty \\
&= \|\lambda(\mathbf{V_D}\boldsymbol{\mu}_D + \mathbf{V}\mathbf{g}) + (1-\lambda)(\mathbf{V_D}\tilde{\boldsymbol{\mu}}_\mathbf{D} + \mathbf{V}\mathbf{g})\|_\infty \\
&\leq \lambda\|(\mathbf{V_D}\boldsymbol{\mu}_D + \mathbf{V}\mathbf{g})\|_\infty + (1-\lambda)\|(\mathbf{V_D}\tilde{\boldsymbol{\mu}}_\mathbf{D} + \mathbf{V}\mathbf{g})\|_\infty \\
&< \lambda + (1-\lambda) = 1,
\end{aligned}$$

where we used property (2.39) and the fact that $(\boldsymbol{\mu}, \boldsymbol{\mu}_D), (\boldsymbol{\mu}, \tilde{\boldsymbol{\mu}}_\mathbf{D})$ are admissible. Therefore, $\lambda\boldsymbol{\mu}_D + (1-\lambda)\tilde{\boldsymbol{\mu}}_\mathbf{D}$ is admissible (notice that the above calculation implies in particular that $\|\mathbf{V}\boldsymbol{\mu} + \mathbf{V_D}(\lambda\boldsymbol{\mu}_D + (1-\lambda)\tilde{\boldsymbol{\mu}}_\mathbf{D})\|_\infty < 1$). □

Proof of Lemma 2.3. Define

$$S_1 = \{\mathbf{g} \in H^1_{\boldsymbol{\mu}} \,:\, \|V_\ell\mathbf{g}(T) + y_\ell\|_\infty \geq 1\},\ S_2 = \{\mathbf{g} \in H^1_{\boldsymbol{\mu}} \,:\, \|V_\ell\mathbf{g}(T) + y_\ell\|_\infty = 1\}.$$

We need to prove that $\inf_{\mathbf{g}\in S_1} \mathcal{I}_{\text{pow}}(\mathbf{g}) = \inf_{\mathbf{g}\in S_2} \mathcal{I}_{\text{pow}}(\mathbf{g})$. Since $S_2 \subseteq S_1$, it follows that $\inf_{g\in S_1} \mathcal{I}_{\text{pow}}(\mathbf{g}) \leq \inf_{g\in S_2} \mathcal{I}_{\text{pow}}(\mathbf{g})$. To prove the reverse inequality, we show that for any $g \in S_1$, there exists $\tilde{\mathbf{g}} \in S_2$ such that $\mathcal{I}_{\text{pow}}(\tilde{\mathbf{g}}) \leq \mathcal{I}_{\text{pow}}(\mathbf{g})$. Pick $\mathbf{g} \in S_1$, and let $t' \in [0,T]$ be the first time such that $|y_\ell + V_\ell\mathbf{g}(t')| = 1$.

Clearly $t' > 0$, since $|y_\ell + V_\ell \mathbf{g}(0)| = |\nu_\ell| < 1$. If $t' = T$, we may take $\tilde{\mathbf{g}}(t) = \mathbf{g}(t)$. If $t' < T$, define $\tilde{\mathbf{g}}(t)$ by time-shifting $\mathbf{g}(T)$ to the right as follows:

$$\tilde{\mathbf{g}}(t) = \begin{cases} \boldsymbol{\mu} & \text{for } 0 \le t < T - t', \\ \mathbf{g}(t - T + t') & \text{for } T - t' \le t \le T. \end{cases}$$

It is easy to check that $\tilde{\mathbf{g}} \in S_2$, and that $\mathcal{I}_{\text{pow}}(\tilde{\mathbf{g}}) \le \mathcal{I}_{\text{pow}}(\mathbf{g})$, because the path $\tilde{\mathbf{g}}$ incurs no cost up to time $T - t'$. Indeed, $\tilde{\mathbf{g}}$ is equal to the constant $\boldsymbol{\mu}$ in the interval $[0, T - t']$, so we have $\mathbf{b}(\tilde{\mathbf{g}}(t)) = \mathbf{b}(\boldsymbol{\mu}) = 0$ and $\tilde{\mathbf{g}}'(t) = 0$, yielding $\int_0^{T-t'} \left(\frac{g_i' - b_i(g_i)}{\gamma_i(g_i)}\right)^2 dt = 0$ and thus

$$\mathcal{I}_{\text{pow}}(\tilde{\mathbf{g}}) = \int_{T-t'}^{T} \left(\frac{\tilde{g}_i' - b_i(\tilde{g}_i)}{\gamma_i(\tilde{g}_i)}\right)^2 dt = \int_0^{t'} \left(\frac{g_i' - b_i(g_i)}{\gamma_i(g_i)}\right)^2 dt \le \mathcal{I}_{\text{pow}}(\mathbf{g}).$$

□

Lemma 2.7. *The function $a \to \psi_\ell^{(a)}$ is non-decreasing for $a > \nu_\ell$ and non-increasing for $a < \nu_\ell$.*

Proof. First suppose $a \ge \tilde{a} > \nu_\ell \ge 0$. The case $a \le \tilde{a} < \nu_\ell \le 0$ is analogous. We want to show that for all $\mathbf{f} \in \mathbf{y} + \mathbf{V} H^1_{\boldsymbol{\mu}}$ such that $f_\ell(T) = a$, there exist a $\tilde{\mathbf{f}} \in \mathbf{y} + \mathbf{V} H^1_{\boldsymbol{\mu}}$ with $\tilde{f}_\ell(T) = \tilde{a}$ and $\mathcal{I}_{\text{cur}}(\tilde{\mathbf{f}}) \le \mathcal{I}_{\text{cur}}(\mathbf{f})$. Since $f_\ell(0) = \nu_\ell < \tilde{a} \le a$ and f is continuous, there exist a $t' \in (0, T)$ such that $f(t') = \tilde{a}$. Define $\tilde{\mathbf{f}}(t)$ as follows:

$$\tilde{\mathbf{f}}(t) = \begin{cases} \boldsymbol{\nu} & \text{for } 0 \le t < T - t' \\ \mathbf{f}(t - T + t') & \text{for } T - t' \le t \le T \end{cases}$$

It is easy to check that $\tilde{\mathbf{f}} \in \mathbf{y} + \mathbf{V} H^1_{\boldsymbol{\mu}}$, $\tilde{f}_\ell(T) = \tilde{a}$ and $\mathcal{I}_{\text{cur}}(\tilde{\mathbf{f}}) \le \mathcal{I}_{\text{cur}}(\mathbf{f})$. The proof that $\psi_\ell^{(a)}$ is non-increasing for $a < \nu_\ell$ goes along the same lines. □

Proof of Proposition 2.1. Following the methods in [192], for $\ell \in \mathcal{E}' = \{\ell \in \mathcal{E} : V_\ell \ne 0\}$ it can be shown that

$$\psi_\ell^{(a)} = \frac{(a - \nu_\ell)^2}{V_\ell \mathbf{M}_T V_\ell^\top}, \tag{2.40}$$

where $\mathbf{M}_t = \boldsymbol{\Gamma}^2 \mathbf{B}^{-1} (\mathbf{I} - e^{-2\mathbf{B}t}) e^{\mathbf{B}(t-T)}$. The corresponding optimal paths for power injections and currents leading to the overload of line ℓ are

$$\begin{aligned} \mathbf{X}^{(\ell)}(t) &= (a - \nu_\ell) \frac{\mathbf{M}_t V_\ell^\top}{V_\ell \mathbf{M}_T V_\ell^\top} + \boldsymbol{\mu} \in \mathbb{R}^{n_w}, \\ \mathbf{Y}^{(\ell)}(t) &= \mathbf{V} \mathbf{X}^{(\ell)}(t) + y \in \mathbb{R}^m. \end{aligned} \tag{2.41}$$

It follows easily that

$$\mathcal{I}^*_{\text{cur}}(\overline{\boldsymbol{\mu}}) = \min_{\ell \in \mathcal{E}'} \frac{(1-|\nu_\ell|)^2}{V_\ell \mathbf{M}_T V_\ell^\top}. \tag{2.42}$$

A straightforward calculation yields the desired result. □

Proof of Lemma 2.4. First notice that a vector $(\boldsymbol{\mu}, \boldsymbol{\mu}_D)$ such that $\|\boldsymbol{\nu}\|_\infty = \|\mathbf{V}\boldsymbol{\mu} + \mathbf{V_D}\boldsymbol{\mu}_D\|_\infty < 1$ is admissible if and only if the following implication holds:

$$\exists \mathbf{g} \in H^1_{\boldsymbol{\mu}} \text{ s.t. } \mathcal{I}_{\text{pow}}(\mathbf{g}) < \varepsilon \log(1/p) \implies \|h^{g,\boldsymbol{\mu},\boldsymbol{\mu}_D}\|_\infty < 1, \tag{2.43}$$

where

$$h^{g,\boldsymbol{\mu},\boldsymbol{\mu}_D}_\ell(t) := \xi_{\boldsymbol{\tau}}(\mathbf{y} + \mathbf{V}\mathbf{g}) = (y_\ell + V_\ell \boldsymbol{\mu})^2 e^{-t/\boldsymbol{\tau}} + \frac{1}{\boldsymbol{\tau}} \int_0^t e^{-(t-s)/\boldsymbol{\tau}} (y_\ell + V_\ell \mathbf{g}(s))^2 ds, \quad \mathbf{y} = \mathbf{V_D}\boldsymbol{\mu}_D.$$

For all $\ell \in \mathcal{E}$ and for all $t \in [0,T]$, $h^{g,\boldsymbol{\mu},\boldsymbol{\mu}_D}_\ell(t)$ is non-negative and convex in $\boldsymbol{\mu}_D$. Using the property in Eq.(2.43), the rest of the proof goes along the lines of the proof of Lemma 2.2. □

Proof of Lemma 2.5. The proof follows easily from the observation that the event $\|\Theta^{\varepsilon,\boldsymbol{\tau}}_\ell\|_\infty \geq 1$ implies the event $\|Y^\varepsilon_\ell\|_\infty \geq \alpha_\ell$. Indeed, it is easy to check that if $|Y^\varepsilon_\ell(t)| < \alpha_\ell$ for all $t \in [0,T]$, then it follow from Eq. (2.5) that $\Theta^{\varepsilon,\boldsymbol{\tau}}_\ell(t) < 1$ for all $t \in [0,1]$. Thus, we have

$$\begin{aligned}
\omega_\ell &= \lim_{\varepsilon \downarrow 0} -\varepsilon \log \mathbb{P}(\|\Theta^{\varepsilon,\boldsymbol{\tau}}_\ell\|_\infty \geq 1) \geq \lim_{\varepsilon \downarrow 0} -\varepsilon \log \mathbb{P}\|Y^\varepsilon_\ell\|_\infty \geq \alpha_\ell \\
&= \inf_{\mathbf{g}:\ \|y_\ell + V_\ell g\|_\infty \geq \alpha_\ell} \mathcal{I}_{\text{pow}}(\mathbf{g}) = \inf_{\mathbf{g}:\ |y_\ell + V_\ell \mathbf{g}(T)| = \alpha_\ell} \mathcal{I}_{\text{pow}}(\mathbf{g}) \\
&= \psi^{(\alpha_\ell)}_\ell \wedge \psi^{(-\alpha_\ell)}_\ell.
\end{aligned}$$

□

Proof of Proposition 2.2. Thanks to Lemma 2.5 we see that $\mathcal{I}^{(LB)}_{\text{tmp},\boldsymbol{\tau}}$ is a lower bound for the temperature decay rate, i.e., $\mathcal{I}^*_{\text{tmp},\boldsymbol{\tau}} \geq \mathcal{I}^{(LB)}_{\text{tmp},\boldsymbol{\tau}}$. Since $\alpha_\ell > 1 > |\nu_\ell|$ $\forall \ell$, Lemma 2.7 implies $\psi^{(\alpha_\ell)}_\ell \wedge \psi^{(-\alpha_\ell)}_\ell \geq \psi^{(1)}_\ell \wedge \psi^{(-1)}_\ell$, yielding $\mathcal{I}^{(LB)}_{\text{tmp},\boldsymbol{\tau}} \geq \mathcal{I}^*_{\text{cur}}$. □

Proof of Proposition 2.3. Thanks to Lemma 2.5 we have $\mathcal{I}^{(LB)}_{\text{tmp},\boldsymbol{\tau}} = \min_{\ell \in \mathcal{E}'} \psi^{(\alpha_\ell)}_\ell \wedge \psi^{(-\alpha_\ell)}_\ell$. From Eq. (2.40) we get $\psi^{(\alpha_\ell)}_\ell = \frac{(\alpha_\ell - \nu_\ell)^2}{V_\ell \mathbf{M}_T V_\ell^T}$ and thus $\psi^{(\alpha_\ell)}_\ell \wedge \psi^{(-\alpha_\ell)}_\ell = \psi^{(\text{sign}(\nu_\ell)\alpha_\ell)}_\ell = \frac{(\alpha_\ell - |\nu_\ell|)^2}{V_\ell \mathbf{M}_T V_\ell^T}$, where $\text{sign}(a) = 1$ if $a \geq 0$ and $\text{sign}(a) = -1$

otherwise, yielding the expression for $\tilde{\mathcal{R}}_{\varepsilon,p}^{(\mathrm{tmp},\boldsymbol{\tau},LB)}$. In the case $\mathbf{B} = b\mathbf{I}$ a straightforward calculation yields the result. □

Proof of Proposition 2.4. In the case $\mathbf{B} = b\mathbf{I}$, according to equation Eq. (2.41), the optimal current paths to overflow in line ℓ and the corresponding decay rate are

$$\begin{aligned} \mathbf{Y}^{(\ell)}(t) &= (\mathrm{sign}(\nu_\ell) - \nu_\ell)\frac{(1-e^{-2bt})e^{b(t-T)}}{1-e^{2bT}}R^\ell + \boldsymbol{\nu}, \\ \psi_\ell &= \frac{b}{1-e^{-2bT}}\frac{(1-|\nu_\ell|)^2}{\sigma_\ell^2}, \end{aligned} \tag{2.44}$$

where $R^\ell := \frac{\mathbf{V}\boldsymbol{\Gamma}^2 V_\ell^T}{V_\ell \boldsymbol{\Gamma}^2 V_\ell^T} \in \mathbb{R}^m$ and $\sigma_\ell^2 = V_\ell \boldsymbol{\Gamma}^2 V_\ell^T$. Take any $\ell^* \in \arg\min_{\ell\in\mathcal{E}'} \psi_\ell$. Recall that ℓ^* depends on the initial condition $\overline{\boldsymbol{\mu}}$, i.e. $\ell^* = \ell^*(\overline{\boldsymbol{\mu}})$. Letting $S^* = \mathrm{sign}(\nu_{\ell^*}) - \nu_{\ell^*} \in \mathbb{R}$ and $R^* = R^{\ell^*}$, the optimal current path to overflow is $\mathbf{f}_*(t) = \mathbf{Y}^{(\ell^*)}(t)$ and in particular $\mathbf{f}_*(0) = \boldsymbol{\nu}, \mathbf{f}_*(T) = S^*R^* + \boldsymbol{\nu}, (\mathbf{f}_*)'(0) = \frac{2be^{bT}}{1-e^{-2bT}}S^*R^*, (\mathbf{f}_*)'(T) = \frac{b(1+e^{-2bT})}{1-e^{-2bT}}S^*R^*$. After a lengthy but straightforward calculation, which is reported below, the formula for the Taylor approximation reads

$$\mathcal{I}_{\mathrm{tmp},\boldsymbol{\tau}}^{(TL)}(\bar{\boldsymbol{\mu}}) = (1+2\tau_0 b)\mathcal{I}_{\mathrm{cur}}^*(\bar{\boldsymbol{\mu}}). \tag{2.45}$$

The capacity region defined by the Taylor approximation is

$$\begin{aligned} &\tilde{\mathcal{R}}_{\varepsilon,p}^{(\mathrm{tmp},\boldsymbol{\tau},TL)} = \\ &= \bigcap_{\ell\in\mathcal{E}'} \{\overline{\boldsymbol{\mu}} \in \mathbb{R}^N : \frac{b(1-|\nu_\ell|)^2}{(1-e^{-2bT})\sigma_\ell^2}(1+2\tau_0 b) > -\varepsilon\log(p)\}, \end{aligned}$$

which can be rewritten as

$$\tilde{\mathcal{R}}_{\varepsilon,p}^{(\mathrm{tmp},\boldsymbol{\tau},TL)} = \bigcap_{\ell\in\mathcal{E}'} \left\{\overline{\boldsymbol{\mu}} \in \mathbb{R}^N \ : \ |\nu_\ell| < 1 - \eta_\ell/\sqrt{1+2\tau_0 b}\right\}.$$

□

Proof of equation (2.29). Since $\xi_{\boldsymbol{\tau}}(\mathbf{f}) = \mathbf{h}$ if and only if $\boldsymbol{\tau}\mathbf{h}' + \mathbf{h} = \mathbf{f}^2$, the temperature rate function reads

$$\mathcal{I}_{\mathrm{tmp},\boldsymbol{\tau}}(\mathbf{h}) = \begin{cases} G(\boldsymbol{\tau},\mathbf{h}) & \text{if } \mathbf{h} \in \xi_{\boldsymbol{\tau}}(\mathbf{y} + \mathbf{V}H^1_{\boldsymbol{\mu}}), \\ \infty & \text{otherwise,} \end{cases}$$

$$G(\boldsymbol{\tau},\mathbf{h}) = \mathcal{I}_{\mathrm{cur}^2}(\boldsymbol{\tau}\mathbf{h}' + \mathbf{h}) = \mathcal{I}_{\mathrm{cur}}(f_{\boldsymbol{\tau}\mathbf{h}'+\mathbf{h}}) = \mathcal{I}_{\mathrm{pow}}(\mathbf{V}^+(f_{\boldsymbol{\tau}\mathbf{h}'+\mathbf{h}} - \mathbf{y})), \tag{2.46}$$

where $\mathcal{I}_{\text{cur}^2}(\mathbf{F}) = \inf_{\mathbf{f}\in H^1_{\boldsymbol{\nu}}:\,\mathbf{f}^2=\mathbf{F}} \mathcal{I}_{\text{cur}}(\mathbf{f})$ is the rate function for the current squared process $(\mathbf{Y}^\varepsilon(t))^2$ and $\mathbf{f}_{\mathbf{F}} := \arg\min_{\substack{\mathbf{f}\in H^1_{\boldsymbol{\nu}}:\,\mathcal{I}_{\text{cur}}(\mathbf{f}),\\ \mathbf{f}^2=\mathbf{F}}}$. Note that $\mathcal{I}_{\text{cur}^2}(\mathbf{F})$ can be written as

$$\mathcal{I}_{\text{cur}^2}(\mathbf{F}) = \sum_{i=1}^{n_w}\int_0^T K_i(\mathbf{F}(t),\mathbf{F}'(t))dt,$$
$$K_i(\mathbf{F}(t),\mathbf{F}'(t)) = \frac{1}{2}\Big[\frac{V_i^+\mathbf{f}'_{\mathbf{F}}(t) - b_i(V_i^+(\mathbf{f}_{\mathbf{F}}(t)-\mathbf{y}))}{\gamma_i(V_i^+(\mathbf{f}_{\mathbf{F}}(t)-\mathbf{y}))}\Big]^2.$$

The partial derivatives of the function

$$\boldsymbol{\tau} \to K_i((\boldsymbol{\tau}\mathbf{h}'+\mathbf{h}),(\boldsymbol{\tau}\mathbf{h}''+\mathbf{h}'))$$

in $\boldsymbol{\tau}=\mathbf{0}$ read

$$\frac{\partial}{\partial\tau_\ell}K_i\Big(\boldsymbol{\tau}\mathbf{h}'+\mathbf{h},\boldsymbol{\tau}\mathbf{h}''+\mathbf{h}'\Big)\Big|_{\boldsymbol{\tau}=\mathbf{0}} = K_i^{(\ell)}(\mathbf{h},\mathbf{h}')\mathbf{h}'_\ell + K_i^{(m+\ell)}(\mathbf{h},\mathbf{h}')\mathbf{h}''_\ell,$$

yielding

$$\sum_{\ell=1}^m \frac{\partial}{\partial\tau_\ell}K_i\Big(\boldsymbol{\tau}\mathbf{h}'+\mathbf{h},\boldsymbol{\tau}\mathbf{h}''+\mathbf{h}'\Big)\Big|_{\boldsymbol{\tau}=\mathbf{0}} = \frac{d}{dt}K_i(\mathbf{h},\mathbf{h}'),$$
$$\sum_{\ell=1}^m \frac{\partial}{\partial\tau_\ell}G(\boldsymbol{\tau},\mathbf{h})|_{\boldsymbol{\tau}=\mathbf{0}} = \sum_{\ell=1}^m \frac{\partial}{\partial\tau_\ell}\mathcal{I}_{\text{cur}^2}\Big(\boldsymbol{\tau}\mathbf{h}'+\mathbf{h}\Big)\Big|_{\boldsymbol{\tau}=\mathbf{0}}$$
$$= \sum_{\ell=1}^m\sum_{i=1}^{n_w}\int_0^T \frac{\partial}{\partial\tau_\ell}K_i\Big(\boldsymbol{\tau}\mathbf{h}'+\mathbf{h},\boldsymbol{\tau}h''+\mathbf{h}'\Big)\Big|_{\boldsymbol{\tau}=\mathbf{0}}$$
$$= \sum_{i=1}^{n_w}\int_0^T \sum_{\ell=1}^m \frac{\partial}{\partial\tau_\ell}K_i\Big(\boldsymbol{\tau}\mathbf{h}'+\mathbf{h},\boldsymbol{\tau}\mathbf{h}''+\mathbf{h}'\Big)\Big|_{\boldsymbol{\tau}=\mathbf{0}}$$
$$= \sum_{i=1}^{n_w}\int_0^T \frac{d}{dt}K_i(\mathbf{h},\mathbf{h}')dt = \sum_{i=1}^{n_w}\Big[K_i(\mathbf{h}(T),\mathbf{h}'(T)) -$$
$$K_i(\mathbf{h}(0),\mathbf{h}'(0))\Big] =: \Phi(\mathbf{f}_{\mathbf{h}}(0),\mathbf{f}_{\mathbf{h}}(T),\mathbf{f}_{\mathbf{h}'}(0),\mathbf{f}_{\mathbf{h}'}(T)) =: \Phi_{\mathbf{f}_{\mathbf{h}}}.$$

If $\boldsymbol{\tau} = \tau_0(1,\dots,1)^T$, $\tau_0 > 0$, we get $\boldsymbol{\tau}\cdot\nabla G(\boldsymbol{\tau},\mathbf{h})|_{\boldsymbol{\tau}=\mathbf{0}} = \tau_0\Phi_{f_h}$. Finally, Eq. (2.28) follows by noticing that if $\mathbf{f}_*$ is the optimal current path and $\mathbf{h}_* = (\mathbf{f}_*)^2$ then $f_{\mathbf{h}_*} = \mathbf{f}_*$. □

Proof of Eqs. (2.34), (2.45). We have

$$K_i(\mathbf{h}_*(T),(\mathbf{h}_*)'(T)) =$$

$$
\begin{aligned}
&= \frac{1}{2\gamma_i^2}\Big(V_i^+\mathbf{f}_*'(T) + bV_i^+(\mathbf{f}_*(T)-\mathbf{y}) - b\mu_i\Big)^2 \\
&= \frac{1}{2\gamma_i^2}\Big(\frac{b(1+e^{-2bT})}{1-e^{-2bT}}S^*V_i^+R + b(V_i^+S^*R^* + V_i^+(\boldsymbol{\nu}-\mathbf{y})) - b\mu_i\Big)^2 \\
&= \frac{\Big(b(1+e^{-2bT})S^*V_i^+R^* + (1-e^{-2bT})b(V_i^+S^*R^*)\Big)^2}{2\gamma_i^2(1-e^{-2bT})^2} \\
&= \frac{2b^2(1-|\nu_{\ell^*}|)^2}{\gamma_i^2(1-e^{-2bT})^2}\Big(V_i^+R^*\Big)^2; \\
&K_i(\mathbf{h}_*(0),(\mathbf{h}_*)'(0)) = \\
&= \frac{1}{2\gamma_i^2}\Big(V_i^+\mathbf{f}_*'(0) + bV_i^+(\mathbf{f}_*(0)-\mathbf{y}) - b\mu_i\Big)^2 \\
&= \frac{1}{2\gamma_i^2}\Big(\frac{2be^{-bT}}{1-e^{-2bT}}S^*V_i^+R^* + bV_i^+(\boldsymbol{\nu}-\mathbf{y}) - b\mu_i\Big)^2 \\
&= \frac{2b^2e^{-2bT}(1-|\nu_{\ell^*}|)^2}{\gamma_i^2(1-e^{-2bT})^2}\Big(V_i^+R^*\Big)^2;
\end{aligned}
$$

$$
\begin{aligned}
\Phi_{\mathbf{f}_*} &= \sum_{i=1}^{n_w}\Big[K_i(\mathbf{h}_*(T),(\mathbf{h}_*)'(T)) - K_i(\mathbf{h}_*(0),(\mathbf{f}_*^2)'(0))\Big] \\
&= \frac{2b^2(1-|\nu_{\ell^*}|)^2}{1-e^{-2bt}}\sum_{i=1}^{n_w}\Big(\frac{V_i^+R^*}{\gamma_I}\Big)^2.
\end{aligned}
$$

The Taylor approximation thus reads

$$
\begin{aligned}
&\mathcal{I}_{\text{tmp},\boldsymbol{\tau}}^{(TL)}(\tau_0,\overline{\boldsymbol{\mu}}) := \mathcal{I}_{\text{cur}}^*(\overline{\boldsymbol{\mu}}) + \tau_0\Phi_h = \\
&\frac{b}{1-e^{-2bT}}\frac{(1-|\nu_{\ell^*}|)^2}{\sigma_{\ell^*}^2} + \frac{2\tau_0b^2(1-|\nu_{\ell^*}|)^2}{1-e^{-2bt}}\sum_{i=1}^{n_w}\Big(\frac{V_i^+R^*}{\gamma_I}\Big)^2 = \\
&\frac{b(1-|\nu_{\ell^*}|)^2}{1-e^{-2bT}}\Big(\frac{1}{\sigma_{\ell^*}^2} + 2\tau_0b\sum_{i=1}^{n_w}\Big(\frac{V_i^+R^*}{\gamma_I}\Big)^2\Big) = \\
&\frac{b(1-|\nu_{\ell^*}|)^2}{(1-e^{-2bT})\sigma_{\ell^*}^2}\Big(1 + 2\tau_0b\sum_{i=1}^{n_w}\frac{\gamma_i^2C_{\ell^*i}^2}{\sigma_{\ell^*}^2}\Big) = \\
&\frac{b(1-|\nu_{\ell^*}|)^2}{(1-e^{-2bT})\sigma_{\ell^*}^2}\Big(1+2\tau_0b\Big) = (1+2\tau_0b)\mathcal{I}_{\text{cur}}^*(\overline{\boldsymbol{\mu}}).
\end{aligned}
$$

□

Chapter 3

Upper Bounds for Line Failure Probability

Contents

In this chapter, we develop upper bounds for line failure probabilities in power grids, under the DC approximation and assuming Gaussian noise for the power injections. Compared to Chapter 2, the results presented in this chapter are not approximations of failure probabilities based on a small-noise asymptotics, but are rigorous upper bounds for the actual failure probability regardless of the magnitude of the noise, and are thus *guaranteed to be conservative*. The bounds are derived using concentration inequalities techniques, and lead to the characterization of safe operational capacity regions that are conservative, convex and polyhedral, making our tools compatible with existing planning methods.

Chapter outline: The chapter is organized as follows. In Section 3.1 we provide a detailed problem formulation, defining the failure probabilities of

interest. Our main results are two different upper bounds on the failure probability that we present in Section 3.2. The first upper bound is explicit, while the second one is sharper and explicit up to a finite-step minimization procedure. These bounds are compared numerically with the exact safe capacity regions in Section 3.3. Proofs are reported in Section 3.4, and concluding remarks are provided in Section 3.5.

3.1 Problem formulation

3.1.1 System model

We model the power grid network as a connected graph $\mathcal{G} = \mathcal{G}(\mathcal{N}, \mathcal{E})$, where $\mathcal{N}$ denotes the set of *buses* and $\mathcal{E}$ the set of directed edges modeling the *transmission lines*. Let $n = |\mathcal{N}|$ be the number of buses and $m = |\mathcal{E}|$ the number of lines. The rest of the model is the same as in Section 1.2, and will not be repeated here. Let $\mathbf{p} \in \mathbb{R}^n$ denote the vector of power injections and $\hat{\mathbf{f}} \in \mathbb{R}^m$ the vector of power flows over the lines. As usual, we use the convention that $p_i \geq 0$ ($p_i < 0$) means that power is generated (consumed, respectively) at bus i.

We make use of the *DC approximation* described in Section 1.2.2, which states that the line power flows $\hat{\mathbf{f}}$ can be written as a linear transformation of the power injections $\mathbf{p}$, i.e.

$$\hat{\mathbf{f}} = \widehat{\mathbf{V}}\mathbf{p}, \tag{3.1}$$

where $\widehat{\mathbf{V}} = \mathbf{D}\mathbf{A}\mathbf{L}^+ \in \mathbb{R}^{m\times n}$ is the PTDF matrix introduced in (1.19) (see Section 1.2.2 for the definition of the matrices $\mathbf{D}$, $\mathbf{A}$ and $\mathbf{L}$).

Transmission lines can fail due to overload. We say that a *line overload* occurs in transmission line ℓ if $|\hat{f}_\ell| > \bar{f}_\ell$, where $\bar{f}_\ell$ is the *line capacity*. If this happens, the line may trip, causing a global redistribution of the line power flows which could trigger cascading failures and blackouts (Section 1.3.1). It is convenient to look at the *normalized line power flow* vector $\mathbf{f} \in \mathbb{R}^m$, defined component-wise as $f_\ell := \hat{f}_\ell/\bar{f}_\ell$ for every $\ell = 1, \dots, m$. The relation between line power flows and normalized power flows can be rewritten as $\mathbf{f} = \mathbf{\Lambda}\hat{\mathbf{f}}$, where $\mathbf{\Lambda} \in \mathbb{R}^{m\times m}$ is the diagonal matrix $\mathbf{\Lambda} := \mathrm{diag}(\bar{f}_1^{-1}, \dots, \bar{f}_m^{-1})$. In view of (3.1), we have

$$\mathbf{f} = \mathbf{V}\mathbf{p}, \tag{3.2}$$

where $\mathbf{V} := \mathbf{\Lambda}\widehat{\mathbf{V}} \in \mathbb{R}^{m\times n}$. Henceforth, we refer to the normalized power flows simply as power flows, unless specified otherwise.

3.1.2 Stochastic power injections and line power flows

In this section we describe our model for the bus power injections. As our focus is on network reliability under uncertainty, we assume that each bus houses a *stochastic power injection* or *load*. This choice allows to model, for example, intermittent power generation by renewable sources or highly variable load.

In order to guarantee that the network balance condition

$$\mathbf{1}^T \mathbf{p} = 0, \tag{3.3}$$

is satisfied even with stochastic inputs, we assume that bus n is a *slack bus*, which means that its power injection is chosen in such a way that the vector of actual power injections is a zero-sum vector as required in (3.3).

More specifically, we assume that the the vector of the first $n-1$ power injections $(p_1, \dots, p_{n-1})$ follows a multivariate Gaussian distribution, with expected value $\boldsymbol{\mu} \in \mathbb{R}^{n-1}$ and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{(n-1)\times(n-1)}$. Since the covariance matrix $\boldsymbol{\Sigma}$ is positive semi-definite, the matrix $\sqrt{\boldsymbol{\Sigma}} \in \mathbb{R}^{(n-1)\times(n-1)}$ is well defined via the Cholesky decomposition of $\boldsymbol{\Sigma}$. We are now able to formally define the vector $\mathbf{p}$ of power injections as the n-dimensional random vector

$$\mathbf{p} = \mathbf{S}(\sqrt{\boldsymbol{\Sigma}}\mathbf{X} + \boldsymbol{\mu}), \tag{3.4}$$

where $\mathbf{X} \sim \mathcal{N}_{n-1}(\mathbf{0}, \mathbf{I}_{n-1})$ is a $(n-1)$-dimensional standard multivariate Gaussian random variable and $\mathbf{S}$ is the matrix

$$\mathbf{S} := \begin{pmatrix} \mathbf{I}_{n-1} \\ -\mathbf{1} \end{pmatrix} \in \mathbb{R}^{n\times(n-1)}.$$

By construction we have $\mathbf{p} = (p_1, \dots, p_{n-1}, -\sum_{i=1}^{n-1} p_i)$, so that (3.3) is satisfied. Note that this formulation allows us to model *deterministic power injections* as well, by means of choosing the corresponding variances and covariances equal to zero (or, from a practical standpoint, equal to very small positive numbers, so that the rank of $\boldsymbol{\Sigma}$ is not affected).

It is well known that an affine transformation of a multivariate Gaussian random variable is again a multivariate Gaussian random variable. Thus, identity (3.4) tells us that the power injections p are indeed Gaussian, and hence, in view of (3.2), so are the line power flows $\mathbf{f}$. As it is convenient to look at the line power flows $\mathbf{f}$ as an affine transformation of *standard independent* Gaussian random variables, combining (3.2) and (3.4), we can write

$$\mathbf{f} = \mathbf{W}\mathbf{X} + \mathbf{W}^{(nom)}\boldsymbol{\mu}, \tag{3.5}$$

where $\mathbf{W} := \mathbf{V}\mathbf{S}\sqrt{\boldsymbol{\Sigma}} \in \mathbb{R}^{m\times(n-1)}$ and $\mathbf{W}^{(nom)} := \mathbf{V}\mathbf{S} \in \mathbb{R}^{m\times(n-1)}$. We denote by $\boldsymbol{\nu} := \mathbf{W}^{(nom)}\boldsymbol{\mu}$ the vector of expected, or nominal, line power flows.

To summarize, the line power flows $\mathbf{f}$ follow a multivariate Gaussian distribution $\mathbf{f} \sim \mathcal{N}_m(\boldsymbol{\nu}, \mathbf{W}\mathbf{W}^T)$, where the network topology and the correlation of the power injections are both encoded in the matrix $\mathbf{W}$. Note in particular that $f_\ell \sim \mathcal{N}(\nu_\ell, \sigma_\ell^2)$, where the variance can be calculated as

$$\sigma_\ell^2 := \sum_{j=1}^{n} W_{\ell,j}^2. \tag{3.6}$$

The main assumption behind our stochastic model is that the power injections are Gaussian. In [16, Section 1.5] it is argued how this assumption, although simplifying, is reasonable in order to model buses that house wind farms. Note that, compared to the power injections model in [16], our formulation allows for general correlations between stochastic injections, as we do not impose any restrictions on the covariance matrix $\boldsymbol{\Sigma}$.

3.1.3 Line failure probabilities

The main goal of this chapter is to understand how the probability of an overload violation depends on the parameters of the systems and characterize which average power injection vectors $\boldsymbol{\mu}$ will make such a probability smaller than a desired target tolerance.

In view of the definition of line overload given in Section 3.1.1, we define the *line failure event* $\mathcal{L}$ as

$$\mathcal{L} := \left\{ \exists\, \ell = 1, \dots, m \ : \ |\hat{f}_\ell| \geq \bar{f}_\ell \right\} = \bigcup_{\ell=1}^{m} \{ |\hat{f}_\ell| \geq \bar{f}_\ell \}.$$

Leveraging the normalized line power flows introduced earlier, we can equivalently rewrite $\mathcal{L}$ as

$$\mathcal{L} = \left\{ \max_{\ell=1,\dots,m} |f_\ell| \geq 1 \right\}.$$

Given a power injection covariance matrix $\boldsymbol{\Sigma}$, define the *risk level* $r(\boldsymbol{\mu})$ associated with a power injection profile $\boldsymbol{\mu}$ as

$$r(\boldsymbol{\mu}) := \mathbb{E}\left[\max_{\ell=1,\dots,m} |f_\ell| \right],$$

which is a well-defined function $r : \mathbb{R}^{n-1} \to \mathbb{R}$ of the average injection vector $\boldsymbol{\mu}$. Indeed, in view of (3.5) we can rewrite $r(\boldsymbol{\mu}) = \mathbb{E}\big[\max_{\ell=1,\dots,m} |W_\ell \mathbf{X} + W_\ell^{(D)} \boldsymbol{\mu}|\big]$, where W_ℓ and $W_\ell^{(D)}$ denote the ℓ-th row of the matrices $\mathbf{W}$ and $\mathbf{W}^{(nom)}$, respectively, and $\mathbf{X} \sim \mathcal{N}_{n-1}(\mathbf{0}, \mathbf{I}_{n-1})$.

We aim to characterize, for a *given* covariance matrix $\mathbf{\Sigma}$, the average power injection vectors $\boldsymbol{\mu}$ that make line failures *rare events*, say $\mathbb{P}(\mathcal{L}) \leq q$ for some very small threshold $q \in (0,1)$ to be set by the network operator. In other words, given $q \in (0,1)$, we aim to determine the region $\mathcal{R}_q^{\text{true}} \subset \mathbb{R}^{n-1}$ defined by

$$\mathcal{R}_q^{\text{true}} := \{\boldsymbol{\mu} \in \mathbb{R}^{n-1} : \mathbb{P}(\mathcal{L}) \leq q\}.$$

Computing the region $\mathcal{R}_q^{\text{true}}$ entails calculating, for every given $\boldsymbol{\mu} \in \mathbb{R}^{n-1}$, the probability $\mathbb{P}_{\boldsymbol{\mu}}(\mathcal{L})$. This, in turn, means solving many high-dimensional integrals containing a multivariate Gaussian density, which is a non-trivial and computationally expensive task (see also the remark in Section 3.3).

To overcome these difficulties, in this chapter we develop analytic tools which are explicit enough to be useful for planning and control of power grids in the short-term. More specifically, in the next section we propose *capacity regions* that can be calculated much faster and that can be used to approximate $\mathcal{R}_q^{\text{true}}$.

3.2 Main results

This section is entirely devoted to the derivation of three new capacity regions $\mathcal{R}_q^{\text{up}}, \mathcal{R}_q^{\star}$, and $\mathcal{R}_q^{\text{c.i.}}$ that we introduce to approximate $\mathcal{R}_q^{\text{true}}$. We first introduce the probabilistic upper bounds on which our method is based in Section 3.2.1, then formally define the regions $\mathcal{R}_q^{\text{up}}, \mathcal{R}_q^{\star}$, and $\mathcal{R}_q^{\text{c.i.}}$ in Section 3.2.2, and lastly discuss the trade-offs between these different regions in Section 3.2.3.

3.2.1 Concentration inequalities

Our methodology relies on a well-known *concentration bound* for a function of Gaussian random variables. Concentration bounds describe the likelihood of a function of many random variables to deviate from its expected value. In our context, we are interested in understanding how likely the random variable $\max_{\ell=1,\dots,m} |f_\ell|$ is to deviate from its expected value $r(\boldsymbol{\mu}) = \mathbb{E}\left[\max_{\ell=1,\dots,m} |f_\ell|\right]$.

Many concentration bounds have been proved in the literature, see [193, Chapter 2] for an overview. The relevant results for our setting are presented and proved later in Section 3.4. The next theorem presents an explicit upper bound for the line failure probability that can be derived using the aforementioned concentration bounds, and which is expressed in terms of $r(\boldsymbol{\mu}) = \mathbb{E}\left[\max_{\ell=1,\dots,m} |f_\ell|\right]$ and the variances $\sigma_1^2, \dots, \sigma_m^2$ of the line power flows.

Theorem 3.1 (Upper bound for line failure probability)**.**
If $r(\boldsymbol{\mu}) < 1$*, then*

$$\mathbb{P}(\mathcal{L}) \leq \exp\Big(-\frac{(1-r(\boldsymbol{\mu}))^2}{2\max_{\ell=1,\dots,m}\sigma_\ell^2}\Big). \tag{3.7}$$

Note that $\mathbb{E}\left[\max_{\ell=1,\dots,m}|f_\ell|\right] = r(\boldsymbol{\mu}) < 1$ is a natural assumption, since the case where $r(\boldsymbol{\mu}) \geq 1$ is definitely not a desirable operational regime for the power grid, since line failures would not be rare events anymore.

3.2.2 Capacity regions

Given $q \in (0,1)$, region $\mathcal{R}_q^{\text{c.i.}}$ is defined as the region that consists of all average power injection vectors $\boldsymbol{\mu}$ such that the upper bound for $\mathbb{P}(\mathcal{L})$ given by the concentration inequality (3.7) is smaller than or equal to q, i.e.

$$\mathcal{R}_q^{\text{c.i.}} := \left\{\boldsymbol{\mu} \in \mathbb{R}^{n-1} \,:\, \exp\Big(-\frac{(1-r(\boldsymbol{\mu}))^2}{2\max_{\ell=1,\dots,m}\sigma_\ell^2}\Big) \leq q\right\},$$

which can be rewritten as

$$\mathcal{R}_q^{\text{c.i.}} = \left\{\boldsymbol{\mu} \in \mathbb{R}^{n-1} \,:\, r(\boldsymbol{\mu}) \leq 1 - \max_{\ell=1,\dots,m}\sigma_\ell\sqrt{2\log q^{-1}}\right\}.$$

Unfortunately, the exact calculation of $r(\boldsymbol{\mu})$ is computationally expensive, for the same reasons as outlined at the end of Section 3.1. Furthermore, we want to have a better analytic understanding of the dependency of $r(\boldsymbol{\mu})$ on the power injection averages $\boldsymbol{\mu}$, on the network topology and on the variances σ_ℓ, something that is hard to obtain from purely numerical procedures. Aiming to overcome these issues, we propose an explicit upper bound for $r(\boldsymbol{\mu})$, namely

$$r(\boldsymbol{\mu}) \leq r^{\text{up}}(\boldsymbol{\mu}) := \max_{\ell=1,\dots,m}|\nu_\ell| + \max_{\ell=1,\dots,m}\sigma_\ell\sqrt{2\log(2m)}, \tag{3.8}$$

where we recall that $\boldsymbol{\nu} = \mathbf{W}^{(nom)}\boldsymbol{\mu}$ is the vector of average line power flows. The bound in (3.8) is proven in Lemma 3.1 and can be used to obtain the following sub-region of $\mathcal{R}_q^{\text{c.i.}}$

$$\mathcal{R}_q^{\text{up}} := \left\{\boldsymbol{\mu} \in \mathbb{R}^{n-1} \,:\, r^{\text{up}}(\boldsymbol{\mu}) \leq 1 - \max_{\ell=1,\dots,m}\sigma_\ell\sqrt{2\log q^{-1}}\right\},$$

which can be rewritten explicitly as

$$\mathcal{R}_q^{\text{up}} = \left\{\boldsymbol{\mu} \in \mathbb{R}^{n-1} \,:\, \max_{\ell=1,\dots,m}|\nu_\ell| \leq 1 - \max_{\ell=1,\dots,m}\sigma_\ell(\sqrt{2\log q^{-1}} + \sqrt{2\log(2m)}\right\}.$$

In terms of $\boldsymbol{\mu}$, we see that $\mathcal{R}_q^{\mathrm{up}}$ is the intersection of half-spaces, and so $\mathcal{R}_q^{\mathrm{up}}$ is convex and polyhedral. A refinement of our analysis (see Lemma 3.1) shows that it is possible to obtain a sharper upper bound $r^\star(\boldsymbol{\mu})$ for $r(\boldsymbol{\mu})$,

$$r(\boldsymbol{\mu}) \leq r^\star(\boldsymbol{\mu}) \leq r^{\mathrm{up}}(\boldsymbol{\mu}),$$

which results in the following region

$$\mathcal{R}_q^\star := \left\{ \boldsymbol{\mu} \in \mathbb{R}^{n-1} \ : \ r^\star(\boldsymbol{\mu}) \leq 1 - \max_{\ell=1,\ldots,m} \sigma_\ell \sqrt{2 \log q^{-1}} \right\}.$$

While there is no analytic expression for $r^\star(\boldsymbol{\mu})$, we show in Section 3.4 that calculating $r^\star(\boldsymbol{\mu})$ requires only the evaluation of a function in a finite number of points, making it a numerically viable approach, and the resulting capacity region remains convex and polyhedral. Summarizing, we have:

Theorem 3.2 (Inclusions among capacity regions)**.** *Given $q \in (0,1)$, if $r(\boldsymbol{\mu}) < 1$, then the following inclusions hold:*

$$\mathcal{R}_q^{\mathrm{up}} \subseteq \mathcal{R}_q^\star \subseteq \mathcal{R}_q^{\mathrm{c.i.}} \subseteq \mathcal{R}_q^{\mathrm{true}}. \tag{3.9}$$

3.2.3 Discussion

We can guarantee that a line overload is a sufficiently rare event by enforcing that the risk level $r(\boldsymbol{\mu})$ is at most $1 - \max_{\ell=1\ldots,m} \sigma_\ell \sqrt{2\log(1/q)}$. This approach has the merit to provide a capacity region $\mathcal{R}_q^{\mathrm{c.i.}}$ that can be expressed as a simple linear condition on the risk level $r(\boldsymbol{\mu})$, but has the drawback that it requires the computation of $r(\boldsymbol{\mu})$, a non-trivial task.

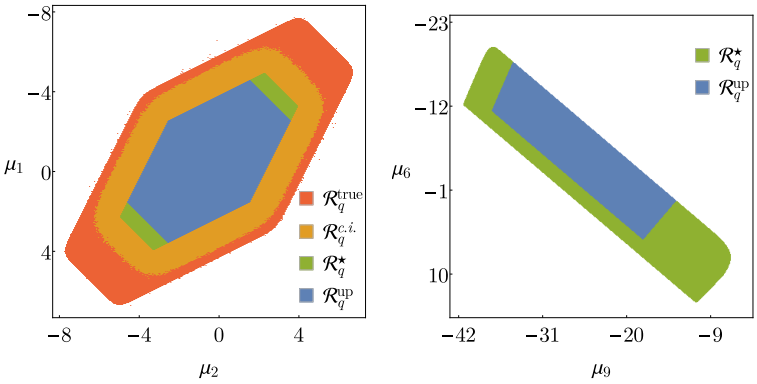
The smaller region $\mathcal{R}_q^{\mathrm{up}}$, although more conservative, is expressed in closed-form and, moreover, its dependency on the parameters $\boldsymbol{\nu}, \sigma$ and m is made explicit. In particular, the maximum standard deviation of the power flows, i.e. $\max_{\ell=1\ldots,m} \sigma_\ell$ plays a big role in defining the capacity regions. Indeed, to larger values of $\max_{\ell=1\ldots,m} \sigma_\ell$ correspond smaller regions, which is intuitive since a bigger variance results in a higher probability of overload.

In between the two regions $\mathcal{R}_q^{\mathrm{up}}$ and $\mathcal{R}_q^{\mathrm{c.i.}}$ lies the intermediate region $\mathcal{R}_q^\star$, which is less conservative than $\mathcal{R}_q^{\mathrm{up}}$ and can be computed very efficiently, even if it cannot be expressed in closed-form (see Section 3.4 for more details). Both regions $\mathcal{R}_q^{\mathrm{up}}$ and $\mathcal{R}_q^\star$ are sufficiently explicit to be used as probabilistic constraints into chance-constrained versions of OPF problems, as studied in [16, 176].

3.3 Numerics

To illustrate how the three new regions compare to $\mathcal{R}_q^{\text{true}}$, we consider first a very simple network with a circuit topology, consisting of 3 buses, all connected with each other by 3 identical lines of unit reactance and capacity $M = 5$. We take the power injections in the non-slack nodes to be independent, zero-mean Gaussian random variables with variance $\varepsilon = 0.5$, which corresponds to taking $\boldsymbol{\mu} = (0, 0)$ and $\Sigma = \varepsilon I_2$. The corresponding four safe capacity regions with $q = 10^{-3}$ are plotted in Fig. 3.1a.

We then plot in Fig. 3.1b the two-dimensional capacity regions $\mathcal{R}_q^{\text{up}}$ and $\mathcal{R}_q^{\star}$ for the IEEE 14-bus test network, where we replace the deterministic power injections at nodes 6 and 9 with Gaussian random variables with average $\boldsymbol{\mu}$ equal to the original deterministic values and variance $\varepsilon = 2 \cdot 10^{-2}$. The line capacities have been chosen to be equal to 1.5 times the average line power flow $\boldsymbol{\nu} = \mathbf{W}^{(nom)}\boldsymbol{\mu}$, and we used $q = 10^{-4}$. The data for $\boldsymbol{\mu}$, line reactances and network topology have been extracted from the MATPOWER package [210]. The regions $\mathcal{R}_q^{\text{c.i.}}$ and $\mathcal{R}_q^{\text{true}}$ have been omitted since the calculation of multivariate Gaussian probabilities such as $\mathbb{P}_{\boldsymbol{\mu}}(\mathcal{L})$ and $r(\boldsymbol{\mu})$ is a non-trivial problem and the subject of active research [51] that goes beyond the scope of this chapter.



(a) 3-bus cycle network

(b) IEEE 14-bus network. The round corners are an artifact of the visualization.

Figure 3.1: Capacity regions comparison for different power grid topologies.

3.4 Mathematical tools

Proposition 3.1 (Unilateral concentration inequality for the maximum of multivariate Gaussian random variables). *Let $\mathbf{X} = (X_1, \dots, X_k) \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ be a multivariate Gaussian random variable, and let $\delta_i := \sqrt{\Sigma_{i,i}}$ be the standard deviation of X_i, $i = 1, \dots, k$. The following concentration inequality holds for every $s \geq 0$:*

$$\mathbb{P}\left(\max_{i=1,\dots,k} |X_i| - \mathbb{E}\left[\max_{i=1,\dots,k} |X_i| \right] \geq s \right) \leq \exp\left(-\frac{s^2}{2 \max_i \delta_i^2} \right).$$

Proof. The multivariate Gaussian vector $\mathbf{X}$ can be seen as an affine transformation $\mathbf{X} = \sqrt{\Sigma}\mathbf{Z} + \boldsymbol{\mu}$ of a standard Gaussian vector $\mathbf{Z} \sim \mathcal{N}_k(0, \mathbf{I}_k)$. Then we apply[193, Theorem 2.26] to the random vector $\mathbf{Z}$ choosing the function $\mathbf{h} : \mathbb{R}^k \to \mathbb{R}$ that maps $\mathbf{Z}$ into $\mathbf{h}(\mathbf{Z}) := \max_{i=1,\dots,k} |(\sqrt{\boldsymbol{\Sigma}})_i \mathbf{Z} + \mu_i|$. A straightforward computation shows that $\mathbf{h}$ is a Lipschitz function with Lipschitz constant equal to $\max_{i=1,\dots,k} \delta_i$. □

Proof of Theorem 3.1. Write

$$\mathbb{P}(\mathcal{L}) = \mathbb{P}\Big(\max_{\ell=1,\dots,m} |f_\ell| - \mathbb{E}\left[\max_{\ell=1,\dots,m} |f_\ell| \right] \geq 1 - \mathbb{E}\left[\max_{i=1,\dots,m} |f_\ell| \right] \Big).$$

Set $s := 1 - \mathbb{E}\left[\max_{\ell=1,\dots,m} |f_\ell|\right] > 0$ and apply Proposition 3.1 to f. Inequality (3.7) follows as the standard deviation of f_ℓ is equal to σ_ℓ, in view of definition (3.6). □

Lemma 3.1 (Upper bounds for the risk level). *Let $r(\boldsymbol{\mu}) := \mathbb{E}\left[\max_{\ell=1,\dots,m} |f_\ell|\right]$, and define*

$$r^\star(\boldsymbol{\mu}) := \inf_{s \in (0,+\infty)} \left\{ \frac{\log(2m)}{s} + \max_{\ell=1,\dots,m} \left(\frac{\sigma_\ell^2}{2} s + |\nu_\ell| \right) \right\}.$$

Then

$$r(\boldsymbol{\mu}) \leq r^\star(\boldsymbol{\mu}) \leq \max_{\ell=1,\dots,m} |\nu_\ell| + \max_{\ell=1,\dots,m} \sigma_\ell \sqrt{2 \log(2m)}. \tag{3.10}$$

Proof. Take $2m$ random variables $Y_1, \dots, Y_{2m}$ defined as

$$Y_j := \begin{cases} f_j & \text{if } j = 1, \dots, m, \\ -f_{j-m} & \text{if } j = m+1, \dots, 2m. \end{cases}$$

From the definition of these random variables it immediately follows that $\max_{\ell=1,\dots,m} |f_\ell| = \max_{j=1,\dots,2m} Y_j$ and therefore $\mathbb{E}\left[\max_{\ell=1,\dots,m} |f_\ell|\right] = \mathbb{E}\left[\max_{j=1,\dots,2m} Y_j\right]$. Note that

$$\lambda_j := \mathbb{E}\left[Y_j\right] = \begin{cases} \nu_j & \text{if } j = 1,\dots,m, \\ -\nu_{j-m} & \text{if } j = m+1,\dots,2m, \end{cases} \tag{3.11}$$

and $\operatorname{Var}\left[Y_j\right] = \operatorname{Var}\left[Y_{j+m}\right] = \sigma_j^2$ for every $j = 1,\dots,m$. For every $j = 1,\dots,2m$, let $m_j(s) := \mathbb{E}\left[\mathrm{e}^{sY_j}\right] = \mathrm{e}^{\sigma_j^2 s^2/2 + \lambda_j s}$ be the moment generating function of the random variable Y_j. Following [52], for any $s \geq 0$ we have

$$\mathrm{e}^{s\,\mathbb{E}\left[\max_{j=1,\dots,2m} Y_j\right]} \leq \mathbb{E}\left[\mathrm{e}^{s \max_{j=1,\dots,2m} Y_j}\right] = \sum_{j=1}^{2m} m_j(s) \leq 2m \max_{j=1,\dots,2m} \mathbb{E}\left[\mathrm{e}^{sY_j}\right].$$

Taking the $\log$ on both sides and rearranging we obtain

$$\begin{aligned} \mathbb{E}\left[\max_{j=1,\dots,2m} Y_j\right] &\leq \inf_{s\in(0,\infty)} \frac{1}{s} \log\left(2m \max_{j=1,\dots,2m} \mathbb{E}\left[\mathrm{e}^{sY_j}\right]\right) \\ &= \inf_{s\in(0,\infty)} \left\{ \frac{\log(2m)}{s} + \frac{1}{s} \log\left[\max_{j=1,\dots,2m} \left(\mathrm{e}^{\sigma_j^2 s^2/2 + \lambda_j s}\right)\right] \right\}, \end{aligned}$$

yielding the first bound, since the RHS is equal to $r^\star(\boldsymbol{\mu})$. If we now define $\hat{\nu} := \max_{\ell=1,\dots,2m} \lambda_\ell = \max_{\ell=1,\dots,m} |\nu_\ell|$ and $\hat{\sigma}^2 = \max_{\ell=1,\dots,m} \sigma_\ell^2$, we have $\max_{j=1,\dots,2m} m_j(s) \leq \mathrm{e}^{\frac{\hat{\sigma}^2}{2} s^2 + \hat{\nu} s}$ for all $s \geq 0$. Thus

$$\mathbb{E}\left[\max_{j=1,\dots,2m} Y_j\right] \leq \frac{\log(2m)}{s} + \frac{\hat{\sigma}^2}{2} s + \hat{\nu}$$

for all $s \in (0, +\infty)$. Optimizing over s in $(0, +\infty)$ yields $\mathbb{E}\left[\max_{j=1,\dots,2m} Y_j\right] \leq \hat{\nu} + \hat{\sigma}\sqrt{2\log(2m)}$, corresponding to $s = \hat{\sigma}^{-1}\sqrt{2\log(2m)}$, thus proving the other inequality in (3.10). □

Lastly, we make some final remarks on how to calculate $r^\star(\boldsymbol{\mu})$ which is the infimum over $(0, \infty)$ of

$$g(s) := \frac{\log(2m)}{s} + \max_{\ell=1,\dots,m} \left(\frac{\sigma_\ell^2}{2} s + |\nu_\ell|\right). \tag{3.12}$$

This can be seen as the point-wise maximum of m functions $g_\ell(s) := \frac{\log(2m)}{s} + \frac{\sigma_\ell^2}{2} s + |\nu_\ell|$, $\ell = 1,\dots,m$. Note that $r^\star(\boldsymbol{\mu})$ can be computed by evaluating the function g in at most $m + m(m-1)/2$ points and then take the minimum value:

the candidate points are the m local minima of the functions $g_1(s), \dots, g_m(s)$ (which are given by $s_\ell^\star := \sqrt{2\log(2m)}/\sigma_\ell$, $\ell = 1, \dots, m$), and the points $s_{i,j} := 2(|\nu_i| - |\nu_j|)/(\sigma_j^2 - \sigma_i^2)$, $\quad i, j = 1, \dots, m,\ i \neq j$, (if they exist and are positive) which are at most $m(m-1)/2$. This analysis implies that the resulting capacity region is convex and polyhedral.

3.5 Concluding remarks

Probabilistic techniques, in particular powerful upper bounds for Gaussian random vectors, can be applied to generate explicit upper bounds for failure probabilities and corresponding safe capacity regions. The resulting regions are convex, polyhedral, and can be efficiently incorporated in optimization routines such as OPF, due to the fact that they can be expressed in closed-form (modulo taking the minimum of a finite set of values as in (3.12)). A limitation of this approach is that it does not scale well with the size of the network. In particular, the upper bounds $r^\star(\boldsymbol{\mu}), r^{\mathrm{up}}(\boldsymbol{\mu})$ are not very tight for larger networks, resulting in over-conservative regions. In order to study a realistic network for the German power grid with approximately 10^4 lines in Chapter 4, we use large deviations theory instead.

The results presented in Chapters 2 and 3 have been focusing on the event a *single line failure*, from the different perspectives of asymptotic approximations (Chapter 2) and non-asymptotic upper bounds (Chapter 3). A natural extension of this line of research is the study of *multiple line failures*, which is carried out in Chapters 4 and 5.

Chapter 4

Emergent Failures and Cascades in Power Grids

Contents

In this chapter, we study in detail the concept of *emergent failures* in power grids, which has been introduced in Section 1.6, and start our analysis of cascading failures. We model power grids as complex networks in which line failures can emerge *indirectly*, or endogenously, as a consequence of stochastic fluctuations of the node inputs (modeling renewable energy production), going beyond traditional models assuming that failures are triggered by deterministic, external events (see Section 1.4.2).

In general, line failures can arise when the network is driven from a stable state to a critically loaded state by external factors. Then, intermittent power generation at the renewable nodes causes random fluctuations in the line power flows, possibly triggering outages and cascading failures. Thus, line failures can emerge indirectly due to the interplay between noisy, correlated power input

at the nodes, the network topology, and power flow physics. We analyze this interplay using large deviations theory. In particular, we can identify the most vulnerable lines, rank them according to their likelihood of failing, and establish the most likely configuration of power inputs leading to failures and subsequent cascades. We find that, when weather correlations are taken into account, an emergent line failure does not occur due to large fluctuations in neighboring nodes only, but as a cumulative effect of small unusual fluctuations in the entire network "summed up" by power flow physics.

Moreover, we show that the most likely configuration of nodal noise responsible for the initial failure impacts the way subsequent failures propagate in the network, leading to a higher number of subsequent failures than one would obtain from traditional failure models, which assume that the initial contingency triggering the cascade is an external event. Finally, our approach sheds insights on why the propagation is often of non-local nature, which is a well-known characteristics of cascading failures in power grids.

Finally, we validate the adequacy of our large deviations framework by showing that both the ranking of lines, the most likely configuration of noise, and the most likely propagation of failures obtained via the large deviations approximation are remarkably close to the exact ones (based on numerical evaluations of pre-limit probabilities). Our analysis is mathematically rigorous and explicit in the small-noise limit $\varepsilon \to 0$, where $\varepsilon > 0$ is related to the forecast error for wind and solar power generation on a selected time window, and, as such, quantifies the magnitude of noise in the system. Our results are validated with data from the German transmission network and several IEEE test cases.

Chapter outline: In the rest of this chapter, we first provide a model description in Section 4.1. Next, we identify and rank the most vulnerable lines in Section 4.2, and we analyze the most likely way for failures to occur in Section 4.3. We extend the analysis to subsequent failures in Section 4.4, and we summarize our results in Section 4.5. Finally, Appendices 4.A, 4.A.4 and 4.B provide details on the mathematical derivation, proofs and data analysis.

4.1 System model

The system model is based on the description in Section 1.2.2, which we briefly recall here. We model a transmission network by a connected graph $\mathcal{G}$ with n nodes representing the *buses*, and m directed edges modeling *transmission lines*. The nominal values of net power injections at the nodes are given by $\boldsymbol{\mu} = \{\mu_i\}_{i=1,\dots,n}$. We model the stochastic fluctuation of the power injections around $\boldsymbol{\mu}$, due to variability in renewable generation, by means of the random vector

$\mathbf{p} = \{p_i\}_{i=1,\dots,n}$, which is assumed to follow a nondegenerate multivariate Gaussian distribution

$$\mathbf{p} \sim \mathcal{N}_n(\boldsymbol{\mu}, \varepsilon\boldsymbol{\Sigma}_p), \tag{4.1}$$

with density

$$\varphi(\mathbf{x}) = \frac{\exp(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T(\varepsilon\boldsymbol{\Sigma}_p)^{-1}(\mathbf{x}-\boldsymbol{\mu}))}{(2\pi)^{\frac{n}{2}}\det(\varepsilon\boldsymbol{\Sigma}_p)^{\frac{1}{2}}}, \tag{4.2}$$

with $\varepsilon\boldsymbol{\Sigma}_p \in \mathbb{R}^{n\times n}$ being the nonsingular covariance matrix of $\mathbf{p}$. We interpret the mean $\boldsymbol{\mu}$ as the nominal, or expected, power generation for the next operational window. In our case study for the German transmission grid, we obtain $\boldsymbol{\mu}$ by solving an OPF problem (see Appendix 1.3.1) based on realistic data for wind and solar generation.

Regarding the terms ε and $\boldsymbol{\Sigma}_p$ in Eq. (4.1), we observe that the distinction between the noise parameter ε and the covariance matrix $\boldsymbol{\Sigma}_p$ is relevant only for the theoretical analysis, where we take the limit $\varepsilon \to 0$ while keeping the matrix $\boldsymbol{\Sigma}_p$ fixed. Indeed, as far as the numerical case study is concerned, all the results are obtained by using the product $\varepsilon\boldsymbol{\Sigma}_p$, which is directly estimated from the SciGRID data. For this reason, without loss of generality, in the case study we normalize $\varepsilon = 1$ and estimate $\boldsymbol{\Sigma}_p$ using ARMA models; for details see Section 4.B.1, which also describes an extension of the model covering both stochastic renewable generators and deterministic conventional power plants (see Section 4.A.1).

The Gaussian assumption is debatable, both for solar and wind power. While such an assumption is consistent with atmospheric physics [16] and recent wind park statistics [106, 12], different models are preferred for different timescales [23, 166, 146, 121]. An extension of our framework to the dynamic model in [121] looks promising, using Freidlin-Wentzell theory as in Chapter 2 of this thesis. For a static non-Gaussian extension, see Appendix 4.A.1.

Assuming the vector $\boldsymbol{\mu}$ has zero sum and using the DC approximation described in Section 1.2.2, which is commonly used in transmission system analysis [151, 175, 202], we can express the line power flows $\mathbf{f} = \{f_i\}_{i=1,\dots,m}$ as

$$\mathbf{f} = \mathbf{V}\mathbf{p}, \tag{4.3}$$

where $\mathbf{V} \in \mathbb{R}^{m\times n}$ is the PTDF matrix (see Section 1.2.2), which encodes the grid topology and parameters. The total net power injected in the network $\sum_{i=1}^{n} p_i$ is non-zero as $\mathbf{p}$ is random. Automated affine response and redispatch mechanisms take care of this issue in power grids. Mathematically, this corresponds to a "distributed slack" in our model: the total power injection mismatch is distributed uniformly among all nodes (the matrix $\mathbf{V}$ accounts for this; see Appendix 4.A.1).

A line *overloads* if the absolute amount of power flowing in it exceeds a given *line limit*. We consider a static setting, and in particular we assume that such overloads immediately lead to the outage of the corresponding line, to which we will henceforth refer simply as *line failure*. The rationale behind this assumption is that there usually are security relays on high-voltage transmission lines performing an emergency shutdown as soon as the current exceeds a dangerous level. Without such mechanisms, lines may overheat, sag, and eventually trip after a variable amount of time.

We can express the line flows in units of the line capacity by incorporating the latter in the definition of $\mathbf{V}$ (see Appendix 4.A.1, as well as Eq. (1.23) in Section 1.2.2), so that $\mathbf{f}$ is the vector of *normalized* line power flows and the failure of line ℓ corresponds to $|f_\ell| \geq 1$. In view of Eqs. (4.2) - (4.3), the line power flows $\mathbf{f}$ also follow a multivariate Gaussian distribution with mean $\boldsymbol{\nu}$ and covariance matrix $\varepsilon\boldsymbol{\Sigma}_f$.

The vector $\boldsymbol{\nu} = \mathbf{V}\boldsymbol{\mu} \in \mathbb{R}^m$ describes the nominal line flows, while the covariance matrix $\varepsilon\boldsymbol{\Sigma}_f = \varepsilon\mathbf{V}\boldsymbol{\Sigma}_p\mathbf{V}^T$ describes the correlations between line flows fluctuations, taking into account both the correlations of the power injections (encoded by $\boldsymbol{\Sigma}_p$) and correlations created by the network topology due to power flow physics (Kirchhoff's laws) via $\mathbf{V}$.

We let the power grid operate on average safely by assuming that $\max_{\ell=1,\dots,m} |\nu_\ell| < 1$, and we consider the small-noise regime $\varepsilon \to 0$, so that only unlikely fluctuations of line flows lead to failures. We are most interested in scenarios where power grids are highly stressed, meaning that the nominal power injections $\{\mu_i\}_{i=1,\dots,n}$ are such that the corresponding nominal line power flows $\{\nu_\ell\}_{\ell=1,\dots,m}$ are close to their thresholds. Such a stress could be caused by very high wind generation [148].

An illustrative scenario is reported in Fig. 4.1, which depicts a snapshot of nominal line flows on the SciGRID German network [118]. SciGRID is a simplified model of the actual German transmission network with $n = 585$ buses and $m = 852$ lines that we use as main illustration. The dataset includes load/generation time series, line limits, grid topology and generation costs. In our case study of the German transmission network, we obtain the nominal vector $\boldsymbol{\mu}$ by solving an OPF problem (see Section 1.3.1). In order to model a heavily-loaded but not overloaded system, in the OPF we scale the line limits by a factor of $\lambda = 0.7$ (see Appendix 4.B.1 for more details).

The fluctuations around the nominal schedule $\boldsymbol{\mu}$ are modeled as a zero-mean Gaussian random vector with covariance matrix $\varepsilon\boldsymbol{\Sigma}_p$, which is estimated using statistical models as explained in Appendix 4.B.1.

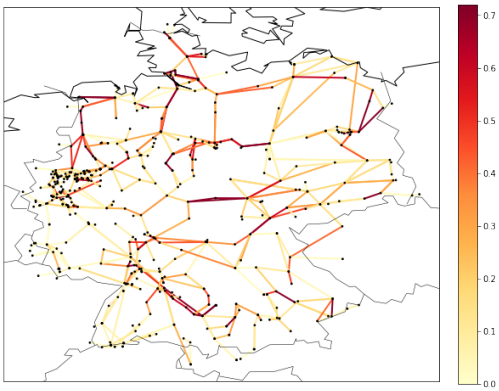


Figure 4.1: Heatmap visualizing the nominal power flows values $|\nu_\ell|$ in the SciGRID model for the snapshot 11am of 01/01/2011.

4.2 Identification and ranking of vulnerable lines

We now turn to the analysis of emergent failures and their propagation using large deviations theory [182]. We begin by deriving the exponential decay of probabilities of single line failure events $|f_\ell| \geq 1$ for $\ell = 1, \dots, m$. As line power flows are Gaussian, we obtain (see Proposition 4.1)

$$I_\ell = -\lim_{\varepsilon \to 0} \varepsilon \log \mathbb{P}_\varepsilon(|f_\ell| \geq 1) = \frac{(1 - |\nu_\ell|)^2}{2\sigma_\ell^2}, \tag{4.4}$$

where $\sigma_\ell^2 = (\mathbf{\Sigma}_f)_{\ell\ell}$. We call I_ℓ the *decay rate* of the failure probability of line ℓ. Thus, for small ε, we approximate the probability of the emergent failure of line ℓ as

$$\mathbb{P}(|f_\ell| \geq 1) \approx \exp(-I_\ell/\varepsilon) = \exp\Big(-\frac{(1 - |\nu_\ell|)^2}{2\varepsilon\sigma_\ell^2}\Big), \tag{4.5}$$

and that of the first emergent failure as

$$\mathbb{P}(\max_\ell |f_\ell| \geq 1) \approx \exp(-\min_\ell I_\ell/\varepsilon). \tag{4.6}$$

These approximations for failure probabilities may not be sharp in general, even when ε is small, since all terms that are decaying subexponentially in $1/\varepsilon$ are ignored. Nevertheless, Eq. (4.5) is quite useful for ranking purposes, allowing to explicitly identify the lines that are most likely to fail. To verify this empirically, we note that the expression in Eq. (4.5) only depends on the product

$\varepsilon\sigma_\ell^2 = \varepsilon(\mathbf{V}\mathbf{\Sigma}_p\mathbf{V}^T)_{\ell\ell}$, and thus, ultimately, only on the product $\varepsilon\mathbf{\Sigma}_p$, which in our case study we estimate directly from the SciGRID data, see Section 4.B.1.

Fig. 4.2 shows the heatmap for the line failure probabilities $\mathbb{P}(|f_\ell| \geq 1)$ (calculated numerically), for the same snapshot as in Fig. 4.1. The most likely line to fail is line 361, which connects two buses housing wind farms (*EON Netz* and *Umspannwerk Kraftwerk Emden*). This line is at capacity ($|\nu_{361}| = 0.7$) and has the highest standard deviation ($\sigma_{361} = 0.142$). However, we observe that a larger $|\nu_\ell|$ does not necessarily imply a higher chance of failure, suggesting that decay rates are a better indicator of system vulnerabilities.

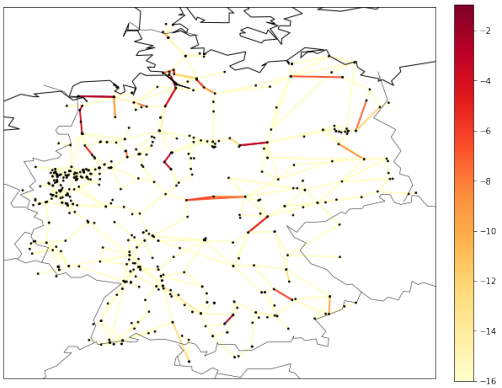


Figure 4.2: Heatmap visualizing the logarithm of the exact overload probabilities $\log_{10}\mathbb{P}(|f_\ell| \geq 1)$ in the SciGRID model, for the snapshot 11am, 01/01/2011.

For instance, several lines in the south of Germany have a moderate to high value $|\nu_\ell|$, see Fig. 4.1. In particular, line 310, which connects buses *Vöhringen Amprion* and *Umspannwerk Dellmensingen*, is at capacity ($|\nu_{310}| = 0.7$), but ranks only 66-th out of 852 lines, with a power flow standard deviation almost one order of magnitude lower than the standard deviation of the most likely line to fail.

Fig. 4.3 depicts the 5% most likely lines to fail, ranked according to the large deviations decay rates $I_\ell = \frac{(1-|\nu_\ell|)^2}{2\sigma_\ell^2}$, where $\sigma_\ell = (\mathbf{\Sigma}_f)_{\ell\ell} = (\mathbf{V}\mathbf{\Sigma}_p\mathbf{V}^\top)_{\ell\ell}$. The ranking based on the large deviations approximation successfully recovers the most likely lines to fail, and, in fact, yields the same ordering as the one based on exact probabilities. As an illustration, in Table 4.1 are reported the indexes, the exact failure probabilities and the decay rates for the 20 most likely lines to fail.

Fig. 4.3 also illustrates the nominal renewable generation mix: the buses housing stochastic power injections have different colors (blue/light blue for

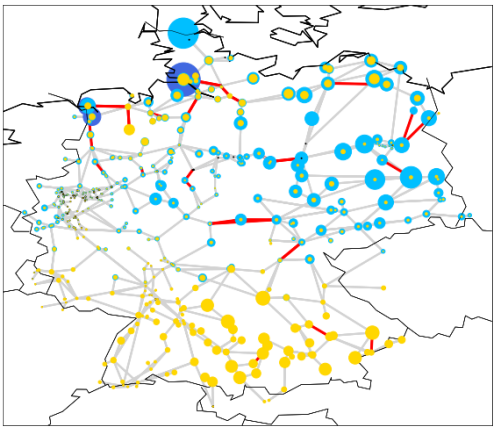


Figure 4.3: Top 5% most likely lines to fail according to Eq. (4.4) for the snapshot 11am, 01/01/2011. The buses housing stochastic power injections have different colors depending on the type of renewable sources (blue for wind offshore, light blue for wind onshore and yellow for solar) and sizes proportional to the absolute values of the corresponding nominal injections.

wind offshore/onshore, yellow for solar) and sizes proportional to the absolute values of the corresponding nominal injections. Many vulnerable lines are located where the most renewable energy production occurs. However, the interplay between network topology, power flows physics and correlation in power injections caused by weather fluctuations, results in a spread-out arrangement of vulnerable lines, which is hard to infer by looking at nominal values only. The large-deviations-based ranking provides a parsimonious way to detect vulnerable lines, and can be used to appreciate qualitative differences among different hours of the day. Table 4.2 lists values for total generation ($\mathcal{G}$), and generation mix for different hours of the day ($p_{\text{w.off}}$ for wind offshore, $p_{\text{w.on}}$ for wind onshore and p_{s} for solar). For example, in the morning there is more solar generation and moderate demand, while in the afternoon there is zero solar generation and higher demand.

In Fig. 4.4 the top 5% most likely lines to fail are depicted (in red) for four different hours of the day, together with the nominal values outputted by the OPF for renewable generation. By comparing Figs. 4.4a-4.4b and Figs. 4.4c-4.4d, for example, we see how solar generation is responsible for an increased number of vulnerable lines in the south of Germany.

ℓ	$\mathbb{P}(\lvert f_\ell \rvert \geq 1)$	I_ℓ
361	1.743e-02	2.225
803	8.228e-04	4.954
19	6.783e-04	5.132
27	6.033e-04	5.240
389	4.503e-04	5.511
390	4.460e-04	5.520
670	3.527e-04	5.737
809	7.575e-05	7.177
586	5.574e-05	7.466
587	5.454e-05	7.486
810	2.496e-05	8.225
712	6.440e-06	9.514
682	5.337e-06	9.693
683	5.318e-06	9.697
714	3.876e-06	9.999
715	1.052e-06	11.249
554	4.267e-07	12.117
488	4.209e-07	12.130
707	1.199e-07	13.341
818	1.199e-07	13.341

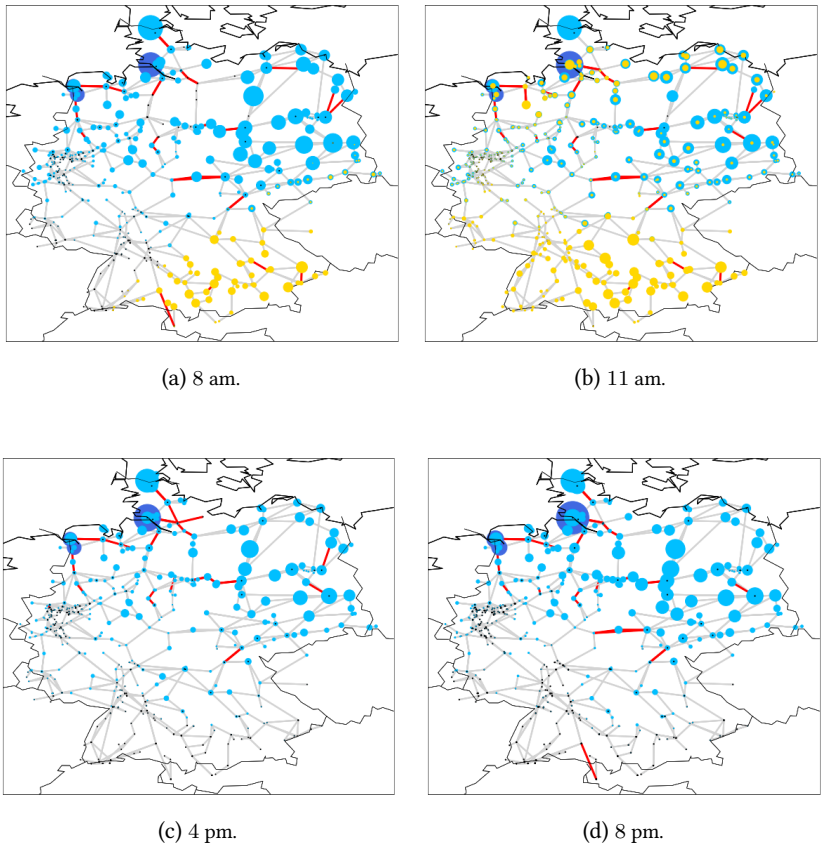
Table 4.1: Line indexes, exact failure probabilities and decay rates of the 20 top most vulnerable lines in the SciGRID model, for the snapshot 11am, 01/01/2011.

hour	G	$p_{\text{w.off}}$	$p_{\text{w.on}}$	p_{sol}
0 am	51.75 GW	1.7 %	35.6 %	0.0%
4 am	44.71 GW	2.0 %	45.6 %	0.0%
8 am	44.83 GW	4.5 %	44.7 %	8.1%
11 am	52.52 GW	4.4 %	32.4 %	17.3%
4 pm	57.56 GW	4.1 %	23.9 %	0.0%
8 pm	54.74 GW	4.1 %	22.9 %	0.0%

Table 4.2: Total generation and renewable percentages for different hours of the day.

4.3 Most likely configuration of power inputs leading to failures

We proceed with an analysis of how emergent failures occur, using again large deviations theory. In particular, we provide an explicit estimate of the most likely power injection that caused a specific emergent failure. To this end, we fix a line ℓ and consider the conditional distribution of $\mathbf{p}$, given $|f_\ell| \geq 1$. The



(a) 8 am. (b) 11 am.

(c) 4 pm. (d) 8 pm.

Figure 4.4: Top 5% most likely lines to fail (in red), together with nominal stochastic generation values, for different hours of the day 01/01/2011. The buses housing stochastic power injections have different colors depending on the type of renewable sources (blue for wind offshore, light blue for wind onshore and yellow for solar) and sizes proportional to the absolute values of the corresponding nominal injections.

mean of this distribution greatly simplifies as $\varepsilon \to 0$ to

$$\mathbf{p}^{(\ell)} = \underset{\mathbf{p}\in\mathbb{R}^n \,:\, |\hat{\mathbf{e}}_\ell^T \mathbf{V}\mathbf{p}| \geq 1}{\arg\inf} \frac{1}{2}(\mathbf{p}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}_p^{-1}(\mathbf{p}-\boldsymbol{\mu}). \tag{4.7}$$

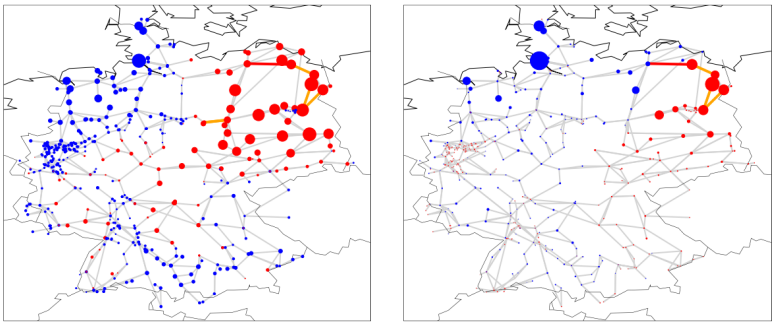
If $\nu_\ell \neq 0$, the solution is unique and reads

$$\mathbf{p}^{(\ell)} = \boldsymbol{\mu} + \frac{(\operatorname{sign}(\nu_\ell) - \nu_\ell)}{\sigma_\ell^2} \boldsymbol{\Sigma}_p \mathbf{V}^T \hat{\mathbf{e}}_\ell, \tag{4.8}$$

where $\mathrm{sign}(a) = 1$ if $a \geq 0$ and -1 otherwise, and $\hat{\mathbf{e}}_\ell \in \mathbb{R}^m$ is the ℓ-th unit vector. As $\varepsilon \to 0$, the conditional variance of $\mathbf{p}$ given $|f_\ell| \geq 1$ decreases to 0 exponentially fast in $1/\varepsilon$, yielding that the conditional distribution of $\mathbf{p}$ given $|f_\ell| \geq 1$ gets sharply concentrated around $\mathbf{p}^{(\ell)}$ (Proposition 4.1 in Appendix 4.A.2). We interpret $\mathbf{p}^{(\ell)}$ as the *most likely* power injection profile, conditional on the failure of line ℓ. The corresponding line power flow profile $\mathbf{f}^{(\ell)} = \mathbf{V}\mathbf{p}^{(\ell)}$ is (see Proposition 4.1)

$$f_k^{(\ell)} = \nu_k + \frac{(\mathrm{sign}(\nu_\ell) - \nu_\ell)}{\sigma_\ell^2} \mathrm{Cov}(f_\ell, f_k), \quad \forall\, k \neq \ell. \tag{4.9}$$

As such, our framework provides more explicit information than the approach in [41], which approximates the most likely way events happen using the mode, without leveraging large deviations. In our validation experiments, we found that the error between $\mathbf{p}^{(\ell)}$ and $\mathbb{E}[\,\mathbf{p}\,|\,|f_\ell| \geq 1]$ is typically less than 1% of the nominal values (see Appendix 4.B.2 for more details).



(a) Non-diagonal $\mathbf{\Sigma}_p$ (b) Diagonal $\mathbf{\Sigma}_p$

Figure 4.5: Representation of the most likely power injection $\mathbf{p}^{(\ell)}$ causing the isolated failure of line 720 (red), and subsequent failures (orange). The bus sizes reflect how much $\mathbf{p}^{(\ell)}$ deviates from $\boldsymbol{\mu}$ (red positive deviations, blue negative). Left, with correlation in noise); Right, without correlation in noise (setting to 0 all the off-diagonals of $\mathbf{\Sigma}_p$). The illustration corresponds to the snapshot 11am, 01/01/2011.

A numerical illustration is given in Fig. 4.5. A key finding is that an emergent line failure does not occur due to large fluctuations only in neighboring nodes, but as a cumulative effect of small unusual fluctuations in the entire network "summed up" by power flow physics, and correlations in renewable energy. Such an emergent failure requires every line flow to be driven to an unusual state $f_k^{(\ell)}$, which deviates from the nominal value ν_k by an amount proportional to the covariance $\mathrm{Cov}(f_\ell, f_k)$, in view of Eq. (4.8).

A non-diagonal noise matrix $\mathbf{\Sigma}_p$ exacerbates these effects. Experiments with our SciGRID case study suggest that, if there is a correlation in noise, for example due to fluctuations in weather patterns, the number of subsequent failures can become higher. Furthermore, it is easier for a failure to be triggered by many small disturbances across the network (Fig. 4.5a), compared to the case where these correlations are not taken into account (Fig. 4.5b). In the latter case, we see a more local effect with relatively larger disturbances.

4.4 Most likely subsequent failures

We continue by investigating the propagation of failures, combining our results describing the most likely power injections configuration leading to the first failure, and the power flow redistribution in the network afterwards. To this end, we first differentiate between different types of line failures, by assessing whether the most likely way for failure of line ℓ to occur is as (i) an *isolated failure*, if $|f_k^{(\ell)}| < 1$ for all line $k \neq \ell$, or (ii) a *joint failure*, if there exists some other line $k \neq \ell$ such that $|f_k^{(\ell)}| \geq 1$.

Any type of line failure(s) cause(s) a global redistribution of the line power flows according to Kirchhoff's laws, which could trigger further outages and cascades. In our setting, the power injections configuration $\mathbf{p}^{(\ell)}$ redistributes across an altered network $\widetilde{\mathcal{G}}^{(\ell)}$ (a subgraph of the original graph $\mathcal{G}$) in which line ℓ (and possible other lines, in case of a joint failure) has been removed, increasing stress on the remaining lines. The way this redistribution happens on $\widetilde{\mathcal{G}}^{(\ell)}$ is governed by power flow physics and we assume that it occurs instantaneously. Extending this to dynamic models [168, 163] is a natural future topic, as transient effects may make the impact of line failures more severe.

The power flow redistribution amounts to compute a new matrix $\widetilde{\mathbf{V}}$ linking the power injections and the new power flows, which can be constructed analogously to $\mathbf{V}$ (see Appendix 4.A.4). The most likely power flow configuration on $\widetilde{\mathcal{G}}^{(\ell)}$ after redistribution is $\widetilde{\mathbf{f}}^{(\ell)} = \widetilde{\mathbf{V}}\mathbf{p}^{(\ell)}$.

In the special case of an isolated failure (say of line ℓ) it is enough to calculate the vector $\boldsymbol{\phi}^{(\ell)} \in \mathbb{R}^{m-1}$ of (normalized) redistribution coefficients, known as *line outage distribution factors* (LODF) [77]. The quantity $\phi_j^{(\ell)}$ takes values in $[-1, 1]$, and $|\phi_j^{(\ell)}|$ represents the percentage of power flowing in line ℓ that is redirected to line j after the failure of the former. The most likely power flow configuration on $\widetilde{\mathcal{G}}^{(\ell)}$ after redistribution then equals $\widetilde{\mathbf{f}}^{(\ell)} = \{f_k^{(\ell)}\}_{k\neq\ell} + f_\ell^{(\ell)}\boldsymbol{\phi}^{(\ell)}$, where $f_\ell^{(\ell)} = \pm 1$ depending on the way the power flow is most likely to exceed the threshold 1. The power flow configuration $\widetilde{\mathbf{f}}^{(\ell)}$ can be efficiently used to determine which lines subsequently fail, by checking for which k we

have $|\widetilde{f}_k^{(\ell)}| \geq 1$, see Proposition 4.3 and Appendix 4.B.2.

There is much evidence that failures propagate non-locally in power grids [99, 103, 109, 160, 116]. To analyze this in our framework we first consider a ring network with $\boldsymbol{\mu} = 0$ and $\boldsymbol{\Sigma}_p = I$. In this network there are two paths along which power can flow between any two nodes, using the convention that a positive flow corresponds to a counter-clockwise direction. If line ℓ fails, the power originally flowing on line ℓ must now flow on the remaining path in the opposite direction. To make this rigorous we show in Lemma 4.1, Appendix 4.A.5 that $\phi_k^{(\ell)} = -1$ for every $k \neq \ell$. As power flows must sum to zero by Kirchhoff's law, neighboring lines tend to have positively correlated power flows, while flows on distant lines exhibit negative correlations. Hence, the power injections that make the power flows in line ℓ exceed the line capacity (say by becoming larger than 1) also make the power flows in the antipodal half of the network negative. These will go beyond the line capacity -1 after the power flow redistributes, see Fig. 4.6.

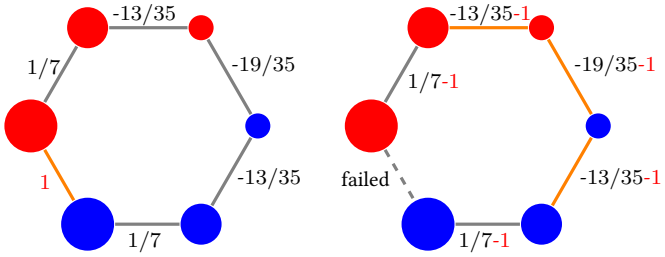


Figure 4.6: Left: most likely power injections $\mathbf{p}^{(\ell)}$ leading to the failure of line ℓ (orange), visualized using the color and size of the nodes (red positive deviations, blue negative), together with power flows $f_k^{(\ell)}$. Right: situation after the power flow redistribution with three subsequent failures and the values $\widetilde{f}_k^{(\ell)} = f_k^{(\ell)} - 1$, $k \neq \ell$.

Fig. 4.7 shows how the emergent isolated failure of line $\ell = 27$ (in red) causes the failure of six more lines $k_1, \ldots, k_6$ (in orange). This example shows how the failure spreads non-locally: in particular, lines 316 and line 602 in the south of Germany are 394 Km and 517 Km far from the original failure. For validation purposes, we found numerically that

$$\mathbb{P}(\text{line } k_j \text{ fails } \forall j = 1, \ldots, 6 \mid |f_{27}| \geq 1) \geq 0.9987.$$

Conversely, the failure of line 27 under the nominal power injection profile leads to only two subsequent failures. The atypical input caused other lines to be more loaded than expected, and these lines get more vulnerable as the cascades progresses, resulting in more subsequent failures.

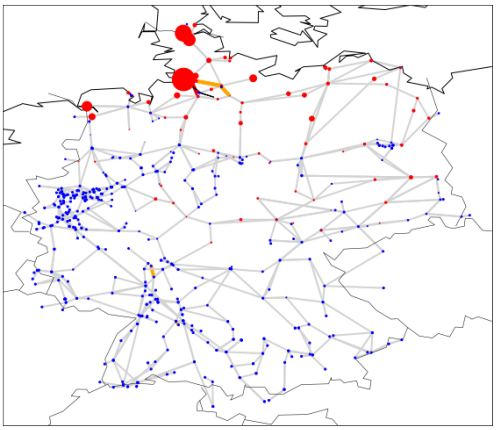


Figure 4.7: The most likely configuration $p^{(27)}$ that leads to the failure of line $\ell = 27$ (in red) in the SciGRID model, for the snapshot 4pm, 01/01/2011. The sizes of the buses reflect how much $p^{(27)}$ deviates from $\boldsymbol{\mu}$ (red for positive deviations, blue for negative ones). The failure of such a line causes, after power redistribution, also the six lines (in orange) to fail.

To validate this insight, we have looked at the first two stages of emergent cascading failures for several IEEE test networks provided in MATPOWER [210], and compare them with those of classical cascading failures, obtained using nominal power injection values rather than the most likely ones and deterministic removal of the initial failing line; see Appendix 4.B.4 for a precise description of the experiment. As before, emergent cascades tend to lead to a higher number of subsequent failures in each stage.

4.5 Concluding remarks

In conclusion, we illustrated the potential of large deviations theory to analyze emergent failures and their propagation in complex networks. Exogenous noise disturbances at the nodes, potentially amplified by correlations, push a complex network into a critical state in which edge failure may emerge. Large deviations theory provides a tool to rank such failures according to their likelihood and predicts how such failures most likely occur and propagate. When an emergent edge failure occurs, its impact on the network can be more significant than a purely exogenous failure, possibly resulting in cascades that propagate quicker than in classical vulnerability analysis.

The accuracy of the small noise limit has been validated in our case study, making the case for applying large deviations techniques to more realistic

models. Finally, in Appendix 4.B.3 we propose an economic application of our approach, showing how our framework can shed light on the trade-off between network reliability and societal costs, expressed in term of energy prices. The impact of uncertainty on energy prices will be investigated further in Chapters 6 and 7 of this thesis. Finally, we note that the analysis carried out in this chapter uses large deviations techniques to study failures and cascades from a microscopic perspective. However, there are phenomena, such as the scale-free nature of blackout sizes (see Section 1.4.2), that are best investigated assuming a macroscopic viewpoint, as we describe in Chapter 5 of this thesis.

Appendix

4.A Model extensions and mathematical results

4.A.1 Detailed model

The model is the one described in Section 1.2.2. For this chapter, we choose to use $\mathbf{L}^+$ rather than $\bar{\mathbf{L}}$ to describe the relationship between the vector net power injections $\mathbf{p} \in \mathbb{R}^n$ and the phase angles $\boldsymbol{\delta} \in \mathbb{R}^n$ they induce in the network nodes:

$$\boldsymbol{\delta} = \mathbf{L}^+ \mathbf{p}. \tag{4.10}$$

This latter identity is particularly useful in our context, since it holds for any vector of power injections $\mathbf{p} \in \mathbb{R}^n$, even if it has no zero sum. Indeed, decomposing the vector p using the basis of eigenvectors $\mathbf{1}, \mathbf{v}_2, \dots, \mathbf{v}_n$ of $\mathbf{L}^+$ one notices that the only component of $\mathbf{p}$ with non-zero sum belongs to the null space of $\mathbf{L}^+$ (generated by the eigenvector $\mathbf{1}$).

This mathematical fact corresponds to the assumption that the power grid has automatic redispatch/balancing mechanisms, in which the total power injection mismatch is distributed uniformly among all the nodes, thus ensuring that the total net power injection is always zero. Conversely, the matrix $\bar{\mathbf{L}}$ does not account for the distributed slack, which needs to be added by post-multiplying by the matrix $\mathbf{S} = \mathbf{I} - \frac{1}{n}\mathbf{J} \in \mathbb{R}^{n \times n}$, where $\mathbf{J} \in \mathbb{R}^{n \times n}$ is the matrix with all entries equal to one.

The real line power flows $\hat{\mathbf{f}}$ are related with the net power injections $\mathbf{p}$ via the linear relationship

$$\hat{\mathbf{f}} = \widehat{\mathbf{V}}\mathbf{p}.$$

As usual, it is convenient to look at the *normalized line power flow* vector $\mathbf{f} \in \mathbb{R}^m$, defined component-wise as $f_\ell = \hat{f}_\ell / M_\ell$ for every $\ell = 1, \dots, m$, where $\bar{f}_\ell$ is the *line capacity* of line ℓ, which is assumed to be given. Line thresholds are in place because a protracted current overload would heat up the line, causing sag,

loss of tensile strength and eventually mechanical failure. If this happens, the failure may cause a global redistribution of the line power flows which could trigger cascading failures and blackouts.

The relation between line power flows and normalized power flows can be rewritten as $\mathbf{f} = \mathbf{\Lambda}\hat{\mathbf{f}}$, where $\mathbf{\Lambda}$ is the $m \times m$ diagonal matrix $\mathbf{\Lambda} = \operatorname{diag}(M_1^{-1}, \ldots, M_m^{-1})$. In view of Eq. (3.1), the normalized power flows $\mathbf{f}$ can be expressed in terms of the power injections $\mathbf{p}$ as $\mathbf{f} = \mathbf{V}\mathbf{p}$, where $\mathbf{V} = \mathbf{\Lambda}\mathbf{D}\mathbf{A}\mathbf{L}^{+} \in \mathbb{R}^{m\times n}$ (see Section 1.2.2). We now briefly outline how the model presented above can be extended to a setting where only a subset of nodes houses stochastic power injections (modeling wind and solar parks), while the other nodes house deterministic injections (corresponding to conventional controllable power plants).

First, we introduce the following notation: if $\mathbf{z}$ is a n-dimensional multivariate Gaussian random vector with mean $\boldsymbol{\lambda}$ and covariance matrix $\mathbf{\Lambda}$, it will be denoted by $\mathbf{z} \sim \mathcal{N}_n(\boldsymbol{\lambda}, \mathbf{\Lambda})$. Define the following:

n_s	number of stochastic buses,
n_d	number of deterministic buses,
$\mathcal{I}_s \subseteq \{1, \ldots, n\}$	indices of stochastic buses,
$\mathcal{I}_d \subseteq \{1, \ldots, n\}$	indices of deterministic buses,
$\mathbf{p}_s = (p_i)_{i\in\mathcal{I}_s} \in \mathbb{R}^{n_s}$	stochastic power injection,
$\mathbf{p}_d = (p_i)_{i\in\mathcal{I}_d} \in \mathbb{R}^{n_d}$	deterministic power injection,
$\mathbf{V}_s \in \mathbb{R}^{m\times n_s}$	matrix consisting of the columns of $\mathbf{V}$ indexed by $\mathcal{I}_s$,
$\mathbf{V}_d \in \mathbb{R}^{m\times n_d}$	matrix consisting of the columns of $\mathbf{V}$ indexed by $\mathcal{I}_d$,
$\mathbf{f}_s = \mathbf{V}_s\mathbf{p}_s \in \mathbb{R}^m$	stochastic component of f,
$\mathbf{f}_d = \mathbf{V}_d\mathbf{p}_d \in \mathbb{R}^m$	deterministic component of f.

If a bus hosts both stochastic and deterministic generators, it is considered a stochastic bus. Stochastic power injections are modelled by means of a n_s-dimensional multivariate Gaussian random vector with mean $\boldsymbol{\mu}_s \in \mathbb{R}^{n_s}$ and covariance matrix $\mathbf{\Sigma}_p \in \mathbb{R}^{n_s\times n_s}$, which we denote by

$$\mathbf{p}_s \sim \mathcal{N}_{n_s}(\boldsymbol{\mu}_s, \varepsilon\mathbf{\Sigma}_p),$$

With the previous notation, the normalized power flows can be decomposed as

$\mathbf{f} = \mathbf{f}_s + \mathbf{f}_d = \mathbf{V}_s \mathbf{p}_s + \mathbf{f}_d$, where

$$\begin{aligned} \mathbf{f}_s &\sim \mathcal{N}_m(\boldsymbol{\nu}_s, \varepsilon \boldsymbol{\Sigma}_f), \\ \boldsymbol{\nu}_s &= \mathbf{V}_s \boldsymbol{\mu}_s, \\ \boldsymbol{\Sigma}_f &= \mathbf{V}_s \boldsymbol{\Sigma}_p \mathbf{V}_s^\top. \end{aligned} \tag{4.11}$$

The nominal power flows values are thus equal to $\boldsymbol{\nu} = \boldsymbol{\nu}_s + \mathbf{f}_d$. The decay rate for an overload in line ℓ, analogously to Eq. (4.8), is given by

$$I_\ell = \inf_{\mathbf{p}_s \in \mathbb{R}^{n_s} \,:\, |\hat{\mathbf{e}}_\ell^\top (\mathbf{V}_s \mathbf{p}_s + \mathbf{f}_d)| \geq 1} \frac{1}{2} (\mathbf{p}_s - \boldsymbol{\mu}_s)^\top \boldsymbol{\Sigma}_p^{-1} (\mathbf{p}_s - \boldsymbol{\mu}_s).$$

Provided that $\nu_\ell \neq 0$, the solution is unique and reads

$$\mathbf{p}_s^{(\ell)} = \frac{(\mathrm{sign}(\nu_\ell) - \nu_\ell)}{\sigma_\ell^2} \boldsymbol{\Sigma}_p \mathbf{V}_s^\top \hat{\mathbf{e}}_\ell + \boldsymbol{\mu}_s \in \mathbb{R}^{n_s}, \tag{4.12}$$

where $\sigma_\ell^2 = (\boldsymbol{\Sigma}_f)_{\ell,\ell}$. The corresponding most likely realization for power flows reads

$$\begin{aligned} \mathbf{f}^{(\ell)} &= \mathbf{V}_s \mathbf{p}_s^{(\ell)} + \mathbf{f}_d \\ &= \frac{(\mathrm{sign}(\nu_\ell) - \nu_\ell)}{\sigma_\ell^2} \mathbf{V}_s \boldsymbol{\Sigma}_p \mathbf{V}_s^\top \hat{\mathbf{e}}_\ell + \boldsymbol{\nu}_s + \mathbf{f}_d \in \mathbb{R}^m. \end{aligned} \tag{4.13}$$

In the next section we prove these claims for the particular case of $n_s = n$.

4.A.2 Large deviation principle in the Gaussian case

In this section we provide proofs for Eqs. (4.4) - (4.9). For the sake of clarity we present here only the proofs for the case $n = n_s$, and we remark that Eqs. (4.12) - (4.13) can be proved along similar lines. In the following, we write $\mathbf{p}_\varepsilon$ and $\mathbf{f}_\varepsilon$ to stress the dependence of the power injections and of the line power flows on the noise parameter ε.

Proposition 4.1. *Assume that* $\max_{j=1,\dots,m} |\nu_j| < 1$. *Then, for every* $\ell = 1, \dots, m$, *the sequence of line power flows* $(\mathbf{f}_\varepsilon)_{\varepsilon > 0}$ *satisfies the large deviation principle*

$$\lim_{\varepsilon \to 0} \varepsilon \log \mathbb{P}(|(\mathbf{f}_\varepsilon)_\ell| \geq 1) = -\frac{(1 - |\nu_\ell|)^2}{2\sigma_\ell^2} = -I_\ell. \tag{4.14}$$

The most likely power injection configuration $\mathbf{p}^{(\ell)} \in \mathbb{R}^n$ *given the event* $|(\mathbf{f}_\varepsilon)_\ell| \geq 1$ *is the solution of the variational problem*

$$\mathbf{p}^{(\ell)} = \operatorname*{arg\,inf}_{\mathbf{p} \in \mathbb{R}^n \,:\, |\hat{\mathbf{e}}_\ell^\top \mathbf{V} \mathbf{p}| \geq 1} \frac{1}{2} (\mathbf{p} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}_p^{-1} (\mathbf{p} - \boldsymbol{\mu}), \tag{4.15}$$

which, when $\nu_\ell \neq 0$, can be explicitly computed as

$$\mathbf{p}^{(\ell)} = \boldsymbol{\mu} + \frac{(\operatorname{sign}(\nu_\ell) - \nu_\ell)}{\sigma_\ell^2} \boldsymbol{\Sigma}_p \mathbf{V}^\top \hat{\mathbf{e}}_\ell.$$

The next proposition shows that the conditional distribution of $\mathbf{p}_\varepsilon$, given $|(\mathbf{f}_\varepsilon)_\ell| \geq 1$, gets concentrated around $\mathbf{p}^{(\ell)}$ exponentially fast as $\varepsilon \to 0$, motivating the interpretation of $\mathbf{p}^{(\ell)}$ as the most likely power injection configuration given the failure of line ℓ.

Proposition 4.2. *Assume that $\max_{k=1,\dots,m} |\nu_k| < 1$, and that $\nu_\ell \neq 0$. Then, for all nodes $i = 1, \dots, n$, and for all $\delta > 0$,*

$$\lim_{\varepsilon \to 0} \varepsilon \log \mathbb{P}((\mathbf{p}_\varepsilon)_i \notin (p_i^{(\ell)} - \delta, p_i^{(\ell)} + \delta) \,\big|\, |(\mathbf{f}_\varepsilon)_\ell| \geq 1) < 0.$$

The line power flows corresponding to the power injection configuration $\mathbf{p}^{(\ell)}$ can be calculated as

$$\mathbf{f}^{(\ell)} = \mathbf{V}\mathbf{p}^{(\ell)} = \boldsymbol{\nu} + \frac{(\operatorname{sign}(\nu_\ell) - \nu_\ell)}{\sigma_\ell^2} \mathbf{V}\boldsymbol{\Sigma}_p \mathbf{V}^\top \hat{\mathbf{e}}_\ell \in \mathbb{R}^m.$$

We observe that the vectors $\mathbf{p}^{(\ell)}$ and $\mathbf{f}^{(\ell)}$ are equal to the conditional expectation of the power injections $\mathbf{p}_\varepsilon$ and power flows $\mathbf{f}_\varepsilon$, respectively, conditional on the failure event $f_\ell = \operatorname{sign}(\nu_\ell)$, namely

$$\begin{aligned} \mathbf{p}^{(\ell)} &= \mathbb{E}[\, \mathbf{p}_\varepsilon \mid (\mathbf{f}_\varepsilon)_\ell = \operatorname{sign}(\nu_\ell)], \\ \mathbf{f}^{(\ell)} &= \mathbb{E}[\, \mathbf{f}_\varepsilon \mid (\mathbf{f}_\varepsilon)_\ell = \operatorname{sign}(\nu_\ell)]. \end{aligned} \tag{4.16}$$

In particular, for every $k = 1, \dots, m$,

$$f_k^{(\ell)} = \nu_k + (\operatorname{sign}(\nu_\ell) - \nu_\ell) \frac{\operatorname{Cov}(f_\ell, f_k)}{\operatorname{Var}(f_\ell)}.$$

Note that the case $\nu_\ell = 0$ has been excluded only for compactness. Indeed, in this special case the variational problem (4.15) has two solutions, $\mathbf{p}^{(\ell,+)}$ and $\mathbf{p}^{(\ell,-)}$. This can be easily explained by observing that if the power flow on line ℓ has mean $\nu_\ell = 0$, then it is equally likely for the overload event $\{|f_\ell| \geq 1\}$ to occur as $\{f_\ell \geq 1\}$ or as $\{f_\ell \leq -1\}$ and the most likely power injection configurations that trigger them can be different.

The previous proposition immediately yields the large deviation principle also for the first line failure event $\|\mathbf{f}_\varepsilon\|_\infty \geq 1$, which reads

$$\lim_{\varepsilon \to 0} \varepsilon \log \mathbb{P}(||\mathbf{f}_\varepsilon||_\infty \geq 1) = - \min_{\ell=1,\dots,m} \frac{(1 - |\nu_\ell|)^2}{2\sigma_\ell^2}.$$

Indeed, the decay rate for the event that at least one line fails is equal to the minimum of the decay rates for the failure of each line. The most likely power injections configuration that leads to the event $\|\mathbf{f}_\varepsilon\|_\infty \geq 1$ is $\mathbf{p}^{(\ell^*)}$ with $\ell^* = \arg\min_{\ell=1,\dots,m} \frac{(1-|\nu_\ell|)^2}{2\sigma_\ell^2}$.

Proof of Proposition 4.1. Let $(\mathbf{Z}^{(i)})_{i\in\mathbb{N}}$ be a sequence of i.i.d. m-dimensional multivariate normal vectors $\mathbf{Z}^{(i)} \sim \mathcal{N}_m(\boldsymbol{\nu}, \boldsymbol{\Sigma}_f)$, and let $\mathbf{S}_k = \frac{1}{k}\sum_{i=1}^k Z^{(i)}$ be the sequence of the partial sums. By setting $\varepsilon = \frac{1}{k}$, it immediately follows that that $\mathbf{f}_\varepsilon \overset{d}{=} S_k$, where $\overset{d}{=}$ denotes equality in distribution. Denote $g(\mathbf{p}) = \frac{1}{2}(\mathbf{p}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}_p^{-1}(\mathbf{p}-\boldsymbol{\mu})$. Following [182, Section 3.D], we get

$$\begin{aligned}
&\lim_{\varepsilon\to 0} \varepsilon \log \mathbb{P}((\mathbf{f}_\varepsilon)_\ell \geq 1) = \lim_{k\to\infty} \frac{1}{k} \log \mathbb{P}((\mathbf{S}_k)_\ell \geq 1) = \\
&= -\inf_{\mathbf{p}\in\mathbb{R}^n \,:\, \hat{\mathbf{e}}_\ell^\top \mathbf{V}\mathbf{p} \geq 1} g(\mathbf{p}) = -\frac{(1-\nu_\ell)^2}{2\sigma_\ell^2}, \qquad (4.17)\\
&\lim_{\varepsilon\to 0} \varepsilon \log \mathbb{P}((\mathbf{f}_\varepsilon)_\ell \leq -1) = \lim_{k\to\infty} \frac{1}{k} \log \mathbb{P}((\mathbf{S}_k)_\ell \leq -1) = \\
&= -\inf_{\mathbf{p}\in\mathbb{R}^n \,:\, \hat{\mathbf{e}}_\ell^\top \mathbf{V}\mathbf{p} \leq -1} g(\mathbf{p}) = -\frac{(-1-\nu_\ell)^2}{2\sigma_\ell^2}. \qquad (4.18)
\end{aligned}$$

The optimizers of problems (4.17) and (4.18) are easily computed respectively as

$$\begin{aligned}
\mathbf{p}^{(\ell,+)} &= \boldsymbol{\mu} + \frac{1-\nu_\ell}{\sigma_\ell^2}\boldsymbol{\Sigma}_p \mathbf{V}^\top \hat{\mathbf{e}}_\ell,\\
\mathbf{p}^{(\ell,-)} &= \boldsymbol{\mu} + \frac{-1-\nu_\ell}{\sigma_\ell^2}\boldsymbol{\Sigma}_p \mathbf{V}^\top \hat{\mathbf{e}}_\ell.
\end{aligned}$$

Note that trivially

$$\inf_{\mathbf{p}\in\mathbb{R}^n \,:\, |\hat{\mathbf{e}}_\ell^\top \mathbf{V}\mathbf{p}| \geq 1} g(\mathbf{p}) = \min\Big\{ \inf_{\mathbf{p}\in\mathbb{R}^n \,:\, \hat{\mathbf{e}}_\ell^\top \mathbf{V}\mathbf{p} \geq 1} g(\mathbf{p}), \inf_{\mathbf{p}\in\mathbb{R}^n \,:\, \hat{\mathbf{e}}_\ell^\top \mathbf{V}\mathbf{p} \leq -1} g(\mathbf{p})\Big\},$$

and thus identities (4.14) and (4.15) immediately follow. □

Proof of Proposition 4.2. We have

$$\begin{aligned}
&\log \mathbb{P}((\mathbf{p}_\varepsilon)_i \notin (p_i^{(\ell)} - \delta, p_i^{(\ell)} + \delta) \,\big|\, |(\mathbf{f}_\varepsilon)_\ell| \geq 1)\\
&= \log \mathbb{P}((\mathbf{p}_\varepsilon)_i \notin (p_i^{(\ell)} - \delta, p_i^{(\ell)} + \delta),\ |(\mathbf{f}_\varepsilon)_\ell| \geq 1) - \log \mathbb{P}(|(\mathbf{f}_\varepsilon)_\ell| \geq 1).
\end{aligned}$$

Denote $g(\mathbf{p}) = \frac{1}{2}(\mathbf{p}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}_p^{-1}(\mathbf{p}-\boldsymbol{\mu})$. From large deviations theory, it holds that that

$$\lim_{\varepsilon\to 0} \varepsilon \log \mathbb{P}(|(\mathbf{f}_\varepsilon)_\ell| \geq 1) = - \inf_{\mathbf{p}\in\mathbb{R}^n \,:\, |\hat{\mathbf{e}}_\ell^\top \mathbf{V} p| \geq 1} g(p) \tag{4.19}$$

$$\begin{aligned}\lim_{\varepsilon\to 0} \varepsilon \log \mathbb{P}((\mathbf{p}_\varepsilon)_i \notin (p_i^{(\ell)} - \delta, p_i^{(\ell)} + \delta),\, |(\mathbf{f}_\varepsilon)_\ell| \geq 1) \\ = - \inf_{\substack{\mathbf{p}\in\mathbb{R}^n \,:\, |\hat{\mathbf{e}}_\ell^\top \mathbf{V} p| \geq 1, \\ |p_i - p_i^{(\ell)}| \geq \delta}} g(p).\end{aligned} \tag{4.20}$$

Define the corresponding decay rates as

$$I_\ell = \inf_{\mathbf{p}\in\mathbb{R}^n \,:\, |\hat{\mathbf{e}}_\ell^\top \mathbf{V}\mathbf{p}| \geq 1} g(\mathbf{p}), \quad J_\ell = \inf_{\substack{\mathbf{p}\in\mathbb{R}^n \,:\, |\hat{\mathbf{e}}_\ell^\top \mathbf{V}\mathbf{p}| \geq 1, \\ |p_i - p_i^{(\ell)}| \geq \delta}} g(\mathbf{p}).$$

Then we can rewrite

$$\lim_{\varepsilon\to 0} \varepsilon \log \mathbb{P}((\mathbf{p}_\varepsilon)_k \notin (p_i^{(\ell)} - \delta, p_i^{(\ell)} + \delta) \,\big|\, |(\mathbf{f}_\varepsilon)_\ell| \geq 1) = -J_\ell + I_\ell,$$

and, therefore, the claim is equivalent to proving that $J_\ell > I_\ell$. Notice that the feasible set of the minimization problem (4.20) is strictly contained in that of the problem (4.19), implying that $J_\ell \geq I_\ell$.

Recall that $\mathbf{p}^{(\ell)}$ is the unique optimal solution of (4.19), and let $\hat{\mathbf{p}}^{(\ell)}$ be an optimal solution of (4.20). Clearly $\hat{p}^{(\ell)}$ is feasible also for problem (4.19). If it was the case that $J_\ell = I_\ell$, then $\hat{\mathbf{p}}^{(\ell)}$ would be an optimal solution for (4.19), and thus by uniqueness ($g(p)$ is strictly convex) $\hat{\mathbf{p}}^{(\ell)} = \mathbf{p}^{(\ell)}$. But this leads to a contradiction, since $\hat{\mathbf{p}}^{(\ell)}$ is by construction such that $|\hat{p}_i - \hat{p}_i^{(\ell)}| \geq \delta$. Hence $J_\ell > I_\ell$ and we conclude that

$$\lim_{\varepsilon\to 0} \varepsilon \log \mathbb{P}((\mathbf{p}_\varepsilon)_i \notin (p_i^{(\ell)} - \delta, p_i^{(\ell)} + \delta) \,\big|\, |(\mathbf{f}_\varepsilon)_\ell| \geq 1) < 0. \quad \square$$

4.A.3 Extension to non-Gaussian case

In this section we briefly describe how to extend the analyis to the non-Gaussian scenario. Consider a model for the power injection vector given by

$$\mathbf{p}_\varepsilon = \boldsymbol{\mu} + \sqrt{\varepsilon}\mathbf{X},$$

where $\boldsymbol{\mu} \in \mathbb{R}^n$ and $\mathbf{X} = (X_1, \ldots, X_n)$ is a random vector with mean 0 and log-moment generating function

$$\log M(\mathbf{s}) = \log \mathbb{E}[e^{\langle \mathbf{s}, \mathbf{X}\rangle}].$$

The power flows vector is thus given by $\mathbf{f}_\varepsilon = \mathbf{V}\,\mathbf{p}_\varepsilon$. Define the Fenchel-Legendre (also known as the convex conjugate) transform of $\log M(\mathbf{s})$, i.e.

$$\Lambda^*(\mathbf{x}) = \sup_{s\in\mathbb{R}^n} \left(\langle \mathbf{s}, \mathbf{x}\rangle - \log M(\mathbf{s})\right).$$

Then, for every $\ell = 1, \dots, m$, the sequence $(\mathbf{f}_\varepsilon)_{\varepsilon>0}$ satisfies the large deviation principle (see [55])

$$\lim_{\varepsilon\to 0} \varepsilon \log \mathbb{P}(|(\mathbf{f}_\varepsilon)_\ell| \geq 1) = - \inf_{\mathbf{x}\in\mathbb{R}^n \,:\, |\hat{\mathbf{e}}_\ell^\top \mathbf{V}(\boldsymbol{\mu}+\mathbf{x})|\geq 1} \Lambda^*(x),$$

and the most likely power injection configuration $\mathbf{p}^{(\ell)} \in \mathbb{R}^n$ given the event $|(\mathbf{f}_\varepsilon)_\ell| \geq 1$ is

$$\mathbf{p}^{(\ell)} = \boldsymbol{\mu} + \operatorname*{arg\,inf}_{\mathbf{x}\in\mathbb{R}^n \,:\, |\hat{\mathbf{e}}_\ell^\top \mathbf{V}(\boldsymbol{\mu}+\mathbf{x})|\geq 1} \Lambda^*(x).$$

The rest of the analysis can then be carried out along similar lines as we did for the Gaussian case.

4.A.4 Power flow redistribution

For every line ℓ define $\mathcal{J}(\ell)$ to be the collection of lines that fail jointly with ℓ as

$$\mathcal{J}(\ell) = \{k \;:\; |f_k^{(\ell)}| \geq 1\}.$$

Let $j(\ell) = |\mathcal{J}(\ell)|$ be its cardinality and note that $j(\ell) \geq 1$ as trivially ℓ always belongs to $\mathcal{J}(\ell)$. Denote by $\widetilde{\mathcal{G}}^{(\ell)}$ the graph obtained from $\mathcal{G}$ by removing all the lines in $\mathcal{J}(\ell)$.

Let us focus first on the case of the isolated failure of line ℓ, that is when $\mathcal{J}(\ell) = \{\ell\}$. In this case $\widetilde{\mathcal{G}}^{(\ell)} = \mathcal{G}(\mathcal{N}, \mathcal{E} \setminus \{\ell\})$ is the graph obtained from $\mathcal{G}$ after removing the line $\ell = (i, j)$. Provided that the power injections remain unchanged, the power flows redistribute among the remaining lines. Using the concept of *effective resistance matrix* $\mathbf{R} \in \mathbb{R}^{n\times n}$ and under the DC approximation, in [40, 165, 171] it is proven that alternative paths for the power to flow from node i to j exist (i.e., $\widetilde{\mathcal{G}}^{(\ell)}$ is still connected) if and only if $x_{i,j}^{-1} R_{i,j} \neq 1$. In other words, $x_{i,j}^{-1} R_{i,j} = 1$ can only occur in the scenario where line $\ell = (i, j)$ is a *bridge*, i.e., its removal results in the disconnection of the original graph $\mathcal{G}$ in two components. If $\widetilde{\mathcal{G}}^{(\ell)}$ is still a connected graph, the power flows after redistribution $\bar{\mathbf{f}}^{(\ell)} \in \mathbb{R}^{m-1}$ are related with the original line flows $\mathbf{f} \in \mathbb{R}^m$ in the network $\mathcal{G}$ by the relation

$$\bar{f}_k^{(\ell)} = f_k + f_\ell^{(\ell)} \phi_k^{(\ell)}, \qquad \text{for every } k \neq \ell,$$

where $f_\ell^{(\ell)} = \pm 1$ depending on the way the power flow on line ℓ exceeded the threshold 1. If $\ell = (i,j)$ and $k = (a,b)$ the coefficient $\phi_k^{(\ell)} \in \mathbb{R}$ can be computed as

$$\phi_k^{(\ell)} = \phi_{(i,j),(a,b)} = x_k^{-1} \cdot \frac{\bar{f}_\ell}{\bar{f}_k} \cdot \frac{R_{a,j} - R_{a,i} + R_{b,i} - R_{b,j}}{2(1 - x_\ell^{-1} R_{i,j})}, \tag{4.21}$$

The ratio $\bar{f}_\ell / \bar{f}_k$ appears in the latter formula since we work with normalized line power flows and we correspondingly defined $\boldsymbol{\phi}^{(\ell)} = \{\phi_k^{(\ell)}\}_{k \neq \ell}$ to be the normalized version of the classical line outage distribution factors (LODF, [77]). Moreover, we define the *most likely* power flows configuration $\widetilde{\mathbf{f}}^{(\ell)} \in \mathbb{R}^{m-1}$ after redistribution as

$$\widetilde{f}_k^{(\ell)} = f_k^{(\ell)} + f_\ell^{(\ell)} \phi_k^{(\ell)}, \quad \text{for every } k \neq \ell. \tag{4.22}$$

4.A.5 Ring topology

We now focus on a particular topology, namely the ring on n nodes, which we use as an illustrative example to show the non-locality of cascades. In this topology, nodes are placed on a ring and each node is connected to its previous and subsequent neighbor. Denote the set of nodes as $\mathcal{N} = \{1, \dots, m\}$ and the set of lines $\mathcal{E} = \{l_1, \dots, l_n\}$, where $l_1 = (n,1), l_2 = (1,2), \dots, l_n = (n-1, n)$. It is easy to prove that, in a ring network with homogeneous line capacities and unitary reactances, $\phi_{\ell,k} = -1$ for every $\ell \neq k$.

Lemma 4.1. *Consider a ring network with homogeneous line capacities ($\bar{f}_\ell = M$ for every line ℓ) and homogeneous unitary reactances ($x_\ell = 1$ for every line ℓ, see Section 1.2.2). Then*

i) The effective resistance between a pair of nodes i, j is given by

$$R_{i,j} = \frac{|j-i|(n - |j-i|)}{n}. \tag{4.23}$$

ii) For every pair of lines $l_k = (k-1, k), l_\ell = (\ell - 1, \ell)$, with $k \neq \ell$, the LODF is constant and equal to $\phi_{(k-1,k),(\ell-1,\ell)} = -1$.

Proof. i) See identity (4) in [8]. ii) First, observe that the effective resistance between two adjacent nodes i and $j = i + 1$ in a circuit graph is equal to $R_{i,j} = \frac{n-1}{n}$, thanks to Eq. (4.23). After a straightforward calculation, and using that $x_\ell^{-1} = 1$, $\bar{f}_\ell = M$ for all lines ℓ, Eq. (4.21) becomes

$$\phi_{(k-1,k),(\ell-1,\ell)} = -\frac{2}{2n\left(1 - \frac{n-1}{n}\right)} = -1. \qquad \square$$

4.A.6 General topology

Going back to the case of a general network topology and any type of failures, isolated or joint, the power flows after redistribution $\bar{\mathbf{f}}^{(\ell)} \in \mathbb{R}^{m-j(\ell)}$ are related with the power injections $\mathbf{p} \in \mathbb{R}^n$ by the relation

$$\bar{\mathbf{f}}^{(\ell)} = \widetilde{\mathbf{V}}^{(\ell)}\mathbf{p},$$

where the $(m - j(\ell)) \times n$ matrix $\widetilde{\mathbf{V}}^{(\ell)}$ can be constructed analogously to $\mathbf{V}$, but considering the altered graph $\widetilde{\mathcal{G}}^{(\ell)}$ instead of $\mathcal{G}$. We define the *most likely* power flow configuration $\widetilde{\mathbf{f}}^{(\ell)}$ after redistribution as

$$\widetilde{\mathbf{f}}^{(\ell)} = \widetilde{\mathbf{V}}^{(\ell)}\mathbf{p}^{(\ell)},$$

which generalizes Eq. (4.22) to any kind of failure, isolated or joint. The next proposition shows that it is enough to look at the vector $\widetilde{\mathbf{f}}^{(\ell)}$ to determine whether a line that survived at the first cascade stage (i.e., that did not fail jointly with ℓ) will fail with high probability or not after the power redistribution (i.e., at the second cascade stage).

Proposition 4.3. *Assume that* $\max_{k=1,\dots,m} |\nu_k| < 1$*, and that* $\nu_\ell \neq 0$*. Then, for all lines* $k \in \mathcal{E} \setminus \mathcal{J}(\ell)$*, and for all* $\delta > 0$*,*

$$\lim_{\varepsilon \to 0} \varepsilon \log \mathbb{P}((\bar{\mathbf{f}}_\varepsilon^{(\ell)})_k \notin (\widetilde{f}_k^{(\ell)} - \delta, \widetilde{f}_k^{(\ell)} + \delta) \,\big|\, |(\mathbf{f}_\varepsilon)_\ell| \geq 1) < 0.$$

In particular, if $|\widetilde{f}_k^{(\ell)}| \geq 1$*, then*

$$\mathbb{P}(|(\bar{\mathbf{f}}_\varepsilon^{(\ell)})_k| \geq 1 \,\big|\, |(\mathbf{f}_\varepsilon)_\ell| \geq 1) \to 1 \quad \text{as } \varepsilon \to 0,$$

exponentially fast in $1/\varepsilon$*.*

Proof. Let $A_{\varepsilon,k}$ denote the event $A_{\varepsilon,k} = \{(\bar{\mathbf{f}}_\varepsilon^{(\ell)})_k \notin (\widetilde{f}_k^{(\ell)} - \delta, \widetilde{f}_k^{(\ell)} + \delta)\}$, and define $Q = \lim_{\varepsilon \to 0} \varepsilon \log \mathbb{P}(A_{\varepsilon,k} \,\big|\, |(\mathbf{f}_\varepsilon)_\ell| \geq 1)$. The proof that $Q < 0$ is analogous to the proof of Proposition 4.2. For the second part, it follows from $\lim_{\varepsilon \to 0} \varepsilon \log \mathbb{P}(A_{\varepsilon,k} \,\big|\, |(\mathbf{f}_\varepsilon)_\ell| \geq 1) = Q$ that for every $\eta > 0$ there exists a $\bar{\varepsilon}$ such that, for every $\varepsilon < \bar{\varepsilon}$,

$$Q - \eta \leq \varepsilon \log \mathbb{P}(A_{\varepsilon,k} \,\big|\, |(\mathbf{f}_\varepsilon)_\ell| \geq 1) \leq Q + \eta,$$

and thus

$$\exp\Big(\frac{Q-\eta}{\varepsilon}\Big) \leq \mathbb{P}(A_{\varepsilon,k} \,\big|\, |(\mathbf{f}_\varepsilon)_\ell| \geq 1) \leq \exp\Big(\frac{Q+\eta}{\varepsilon}\Big).$$

Let $A^c_{\varepsilon,k}$ denote the complementary event of $A_{\varepsilon,k}$. Since $|\widetilde{f}^{(\ell)}_k| \geq 1$, for δ sufficiently small we have

$$A^c_{\varepsilon,k} = \{|(\bar{\mathbf{f}}^{(\ell)}_\varepsilon)_k - \widetilde{f}^{(\ell)}_k| < \delta\} \subseteq \{|(\bar{\mathbf{f}}^{(\ell)}_\varepsilon)_k| \geq 1\},$$

yielding

$$\mathbb{P}(|(\bar{\mathbf{f}}^{(\ell)}_\varepsilon)_k| \geq 1 \,\big|\, |(\mathbf{f}_\varepsilon)_\ell| \geq 1) \geq \mathbb{P}(A^c_{\varepsilon,k} \,\big|\, |(\mathbf{f}_\varepsilon)_\ell| \geq 1) \geq 1 - \exp\Big(\frac{Q+\eta}{\varepsilon}\Big).$$

Since $Q < 0$, the result follows. □

4.B Numerical case study

4.B.1 SciGRID German network

We now demonstrate our methodology in the case of a real-world power grid and a realistic system state.

Dataset Description

We perform our experiments using PyPSA, a free software toolbox for power system analysis [26]. We use the dataset described in [24, 25], which provides a model of the German electricity system based on SciGRID and OpenStreetMap [118, 139]. The dataset includes load/generation time series and geographical locations of the nodes, differentiating between renewable and conventional generation. It also provides data for transmission lines limits, transformers, generation capacity and marginal costs, allowing us to couple our theoretical analysis with realistic OPF computations. The time-series provide hourly data for the entire year 2011. For more technical information on the dataset, we refer to [24, 25].

The SciGRID German network consists of 585 buses, 1423 generators including conventional power plants and wind and solar parks, 38 pump storage units, 852 lines and 96 transformers. For the analysis carried out in this chapter, storage units are not included and we exclude transformer failures. The renewable generators are divided in three classes, solar, wind onshore and wind offshore. Each bus can house multiple generators, both renewable and conventional, but it is limited to at most one renewable generator for each class. Let $\mathcal{N}_{\text{w.off}}, \mathcal{N}_{\text{w.on}}, \mathcal{N}_{\text{sol}}$ denote the set of buses housing, respectively, wind offshore, wind onshore and solar generators, with $|\mathcal{N}_{\text{w.off}}| = 5$, $|\mathcal{N}_{\text{w.on}}| = 488$, $|N_{\text{sol}}| = 489$, and $\mathcal{N}_{\text{w.off}} \subseteq \mathcal{N}_{\text{w.on}} \subseteq \mathcal{N}_{\text{sol}}$. The remaining 96 buses house 441 conventional generators.

Let $n_s = 489$ denote the total number of buses housing renewable generators. If a bus houses both renewable and conventional generators, it will be considered a *stochastic* bus for our decomposition formulation. We model stochastic net power injections by means of a multivariate Gaussian random vector $\mathbf{p}_s \sim \mathcal{N}_{n_s}(\boldsymbol{\mu}_s, \varepsilon\boldsymbol{\Sigma}_p)$.

As explained above, the distinction between the noise parameter ε and the covariance matrix $\boldsymbol{\Sigma}_p$ is relevant only for the theoretical analysis. As a consequence, in the following we will take $\varepsilon = 1$ and refer to the covariance matrix of $\mathbf{p}_s$ simply as $\boldsymbol{\Sigma}_p$.

Data-based model for $\boldsymbol{\mu}_s$

In order to get a realistic nominal line flows value $\boldsymbol{\nu}$, we perform a linear OPF relative to the day $01/01/2011$, for different hours of the day. A linear OPF consists of minimizing the total cost of generation, subject to energy balance, generation and transmission lines constraints, under the assumptions of the DC approximations. In order to model a heavily-loaded but not overloaded system, in the OPF we scale the true line limits $\bar{f}_\ell$ by a contingency factor of $\lambda = 0.7$. This is a common practice in power engineering that allows room for reactive power flows and stability reserve.

More precisely, let $g(i)$ be the generation at bus i as outputted by the OPF for a given hour, and let us write it as $g(i) = g_r(i) + g_d(i)$, with $g_r(i)$ the power produced by renewable generators attached to the bus, and $g_d(i)$ the power supplied by conventional generators. If the demand at bus i is given by $d(i)$, then the average stochastic power injection vector $\boldsymbol{\mu}_s \in \mathbb{R}^{n_s}$ is modeled as

$$(\boldsymbol{\mu}_s)_i = g_r(i) + g_d(i) - d(i), \quad i = 1, \dots, n_s,$$

while the deterministic power injection reads $p_i = g_d(i) - d(i)$ for $i \in \mathcal{I}_d$.

Data-based model for $\boldsymbol{\Sigma}_p$

In order to model the fluctuations of renewable generation around the nominal values, and thus estimate $\boldsymbol{\Sigma}_p$, we use realistic hourly values of wind and solar energy production to fit a stochastic model. We then use the steady-state covariance of the model residuals as an estimate for $\boldsymbol{\Sigma}_p$. Following [122], we choose to use AutoRegressive-Moving-Average (ARMA) models, which we describe in detail below. Note that we do not aim to find the best possible stochastic model for renewable generation, which is beyond the scope of this chapter, but instead to provide an estimate for the covariance matrix $\boldsymbol{\Sigma}_p$ in order to validate our theoretical results, which are asymptotically valid in a small-noise regime. We speculate that more sophisticated models, and/or data

on smaller time-scales, may lead to smaller values for the correlations in $\mathbf{\Sigma}_p$, thus getting closer to the small noise limit.

We now describe the estimation procedure for $\mathbf{\Sigma}_p$ (as mentioned before we normalize $\varepsilon = 1$ in our empirical study). The SciGRID dataset contains time series

$$\mathbf{y}_{\text{w.off}} \in \mathbb{R}^{M\times 5}, \mathbf{y}_{\text{w.on}} \in \mathbb{R}^{M\times 488}, \mathbf{y}_{\text{sol}} \in \mathbb{R}^{M\times 489},$$

for the available power output of wind offshore, wind onshore and solar generators, for each hour of the year 2011, accounting for a total of $M = 8760$ measurements for each generator [24]. For each time series, $y_{(\cdot)}(t,j)$ denote the available power output at time t for the j-th generator of a given type, in MW units.

Wind power model

As a pre-processing step, we merge together the two time series $\mathbf{y}_{\text{w.off}}, \mathbf{y}_{\text{w.on}}$ by summing up the onshore and offshore wind power at the buses $\mathcal{N}_{\text{w.off}} \subseteq \mathcal{N}_{\text{w.on}}$. This yields the time series of wind power production

$$y_{\text{w}}(t,j) = y_{\text{w.on}}(t,j) + \mathbb{1}_{j\in\mathcal{N}_{\text{w.off}}} y_{\text{w.off}}(t,j),$$

where $\mathbb{1}_{\{\}}$ is the indicator function of the event in the bracket, taking value 1 if the event is satisfied, and 0, otherwise.

We select one portion of the data $\{1,\dots,T\} \subseteq \{1,\dots,M\}$, corresponding to the month of January, to be used to fit the model. For each windpark j, following [122] we consider an ARMA(1,24) model of the form

$$\begin{aligned} x(t,j) =& a_{1,j}x(t-1,j) + e(t,j) \\ &+ m_{1,j}e(t-1,j) + \ldots + m_{24,j}e(t-24,j), \end{aligned}$$

where $x(t-1,j)$ is the auto-regressive term, and $e(t-k,j)$, $k = 1,\dots,24$, are the white-noise error terms. For each windpark j, we fit the above model to the wind power data $\{y_{\text{w}}(t,j)\}_{t=1:T}$ in R using the function *arima*, and consider the time series of the residuals $e_{\text{w}}(1,j),\dots,e_{\text{w}}(T,j)$.

The empirical variance of the residuals is used as proxy for the variance of the output of windpark j, namely

$$(\mathbf{\Sigma}_{\text{w}})_{jj} = \widehat{\text{Var}}(e_{\text{w}}(1,j),\dots,e_{\text{w}}(T,j)),$$

where $\widehat{\text{Var}}$ denotes the empirical variance. In a similar way, the empirical covariance of the residuals is used to model the covariance between the output of windparks i and j, $i \neq j$, namely

$$(\mathbf{\Sigma}_{\text{w}})_{ij} = \widehat{\text{Cov}}\Big(\{e_{\text{w}}(t,i)\}_{t=1:T}, \{e_{\text{w}}(t,j)\}_{t=1:T}\Big),$$

where $\widehat{\mathrm{Cov}}$ denotes the empirical covariance.

Solar power model

State-of-the-art models for solar irradiance often combine statistical techniques with cloud motion analysis and numerical weather prediction (NWP) models, see [4] for a review. Since the available data in our case study are limited to historical records for power production of solar generators, and do not include any weather data, we used the purely statistical model ARMA(*p*,*q*), which has been used succesfully in [93].

Regarding the orders p, q of the ARMA model, after some exploratory analysis we decided to use an ARMA(24,24) model with all parameters fixed to 0, except for the ones corresponding to the seven hours before, and the one corresponding to twenty-four hours before. The rationale behind this choice is that by using the value corresponding to twenty-four hours before, we capture the dependency on the hour of the day, while the values from 7 hours before capture the shape of the current day.

More precisely, the model reads

$$\begin{aligned} x(t,j) = {} & a_{1,j}x(t-1,j) + \ldots + a_{7,j}x(t-7,j) \\ & + a_{24,j}x(t-24,j) \\ & + e(t,j) + m_{1,j}e(t-1,j) + \ldots + m_{7,j}e(t-7,j) \\ & + m_{24,j}e(t-24,j). \end{aligned}$$

For each solar park j, we fit the above model to the solar power data $(y_{\mathrm{sol}}(t,j))_{t=1:T}$, using again the R function *arima*, and consider the time series of the residuals $(e_{\mathrm{sol}(t,j)})_{t\in\mathcal{D}}$, where $\mathcal{D} \subseteq \{1,\ldots,T\}$ denotes the set of daylight hours of January 2011. The covariance matrix for the solar power generation is obtained as

$$(\mathbf{\Sigma}_{\mathrm{sol}})_{ij} = \widehat{\mathrm{Cov}}\Big((e_{\mathrm{sol}}(t,i))_{t\in\mathcal{D}}, (e_{\mathrm{sol}}(t,j))_{t\in\mathcal{D}}\Big).$$

Since we perform numerical experiments for different hours of the day 01/01/2011, we need to model renewable fluctuations taking into account whether or not we consider a daylight hour, as there is no solar energy production before sunrise and after sunset. In view of this, and assuming that the residuals for the wind and solar models are independent (see [208]), we model the covariance matrix relative to an hour h as

$$\mathbf{\Sigma}_p(h) = \mathbf{\Sigma}_{\mathrm{w}} + \mathbb{1}_{h\in\mathcal{D}_1}\mathbf{\Sigma}_{\mathrm{sol}},$$

where $\mathcal{D}_1 \subseteq \{1,\ldots,24\}$ denotes the set of daylight hours of 01/01/2011.

The magnitude of power injections noise at bus i is quantified by the standard deviation $\sqrt{(\mathbf{\Sigma}_p)_{ii}}$, expressed as a percentage of the combined installed capacity of wind and solar generators located at bus i.[1] In our numerical study, we find that for daylight hours the mean of these standard deviations across all buses is 8.5%, while during nighttime the mean reduces to 5%.

Data-based model for $\mathbf{\Sigma}_f$

In view of Eq. (4.11), the covariance matrix for the line power flows $\mathbf{f}_s$ is calculated as $\mathbf{\Sigma}_f = \mathbf{V}_s \mathbf{\Sigma}_p \mathbf{V}_s^\top$. The magnitude of power flows noise is quantified by the standard deviations $\sigma_\ell = \sqrt{(\mathbf{\Sigma}_f)_{\ell\ell}}$. Since the nominal values for the power flows ν_ℓ have been standardized as fractions of line capacities, and thus range within the interval $[-1, 1]$, the values of σ_ℓ describe the magnitude of the power flows noise as a percentage of the corresponding line capacity. In our numerical study, we find that for daylight hours the power flow standard deviations lie within the range $[0.00007, 0.14219]$, with mean 0.0228, while during nighttime the range is $[0.00001, 0.14203]$, with mean 0.0131.

4.B.2 German network: Most likely power injections

In order to keep the notation light, in the following two subsections we omit the subscript s (which refers to stochastic power injections) from the vectors $\boldsymbol{\mu}$, $\mathbf{p}^{(\ell)}$, $\mathbf{p}_\varepsilon$, $\bar{\mathbf{p}}_\varepsilon^{(\ell)}$.

The small-noise regime theoretical power injections configuration responsible for the failure of line ℓ, as given by Eq. (4.16), reads $\mathbf{p}^{(\ell)} = \mathbb{E}\,[\,\mathbf{p}_\varepsilon \,|\, (\mathbf{f}_\varepsilon)_\ell = \operatorname{sign}(\nu_\ell)]$. As an illustration, Fig. 4.5 depicts $\mathbf{p}^{(\ell)}$ leading to the isolated failure of line 720. The bus sizes reflect how much $\mathbf{p}^{(\ell)}$ deviates from $\boldsymbol{\mu}$, and the color-coding uses red for positive deviations, blue for negative ones.

In order to validate the accuracy of the large-deviations approach, we compare $\mathbf{p}^{(\ell)}$ to the pre-limit conditional expectation of power injections given the failure of line ℓ, namely

$$\bar{\mathbf{p}}_\varepsilon^{(\ell)} = \mathbb{E}\,[\,\mathbf{p}_\varepsilon \,|\, |(\mathbf{f}_\varepsilon)_\ell| \geq 1],$$

which according to Proposition 4.2 converges to $\mathbf{p}^{(\ell)}$ in the limit as $\varepsilon \to 0$. As a measure of error, we consider, for each line ℓ,

$$\operatorname{err}(\ell) = \frac{1}{n_s} \sum_{i=1}^{n_s} \Big| \frac{(\mathbf{p}^{(\ell)})_i - (\bar{\mathbf{p}}_\varepsilon^{(\ell)})_i}{\mu_i} \Big|,$$

[1] We note that normalizing the error using the installed capacity of a generator is standard in the literature [89].

which quantifies the difference between $\mathbf{p}^{(\ell)}$ and $\bar{\mathbf{p}}_\varepsilon^{(\ell)}$, expressed as a percentage of the nominal values $\boldsymbol{\mu}$, averaged across all stochastic nodes. We found that, for the same hour as in Fig. 4.5, the average error across all lines is $\widehat{\mathrm{err}} = \frac{1}{m}\sum_{\ell=1}^{m} \mathrm{err}(\ell) = 0.2\%$, with a maximum value of 2.6%, see Fig. 4.8. Table 4.3 shows that the errors are uniformly small across different hours.

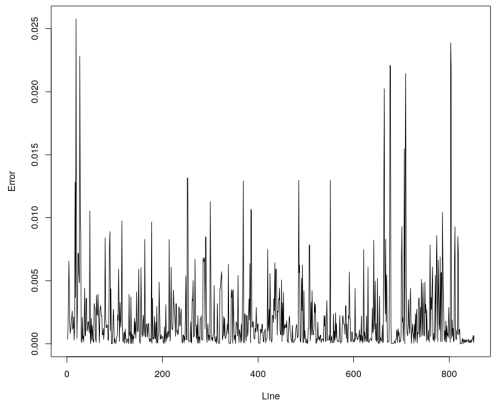


Figure 4.8: Relative error $\mathrm{err}(\ell)$ at 11 am, for $\ell = 1, \ldots, 852$.

hour	$\widehat{\mathrm{err}}$	$\max \mathrm{err}(\ell)$
4 am	0.1%	1.5%
8 am	0.4%	4.6%
11 am	0.2%	2.6%
4 pm	0.1%	2.3%

Table 4.3: Average and maximum $\mathrm{err}(\ell)$ for different hours.

Failure propagation

In view of Proposition 4.3, the subsequent six failures have been determined by looking at the vector $\widetilde{\mathbf{f}}^{(\ell)} = \widetilde{\mathbf{V}}^{(\ell)}\mathbf{p}^{(\ell)}$, and checking whether $|\widetilde{f}_k^{(\ell)}| \geq 1$ for each line $k \neq \ell$. According to Proposition 4.3, the pre-limit conditional probabilities

$$\mathbb{P}(|(\widetilde{\mathbf{f}}_\varepsilon^{(\ell)})_{k_j}| \geq 1 \,|\, |(\mathbf{f}_\varepsilon)_\ell| \geq 1)$$

converge exponentially fast to 1 as $\varepsilon \to 0$, and in particular the cumulative distribution functions

$$\mathbb{P}((\bar{\mathbf{f}}_\varepsilon^{(\ell)})_{k_j} \leq x \,|\, |(\mathbf{f}_\varepsilon)_\ell| \geq 1)$$

converge to the deterministic distribution $\widetilde{f}_{k_j}^{(\ell)}$. In order to validate our methodology, we numerically evaluate

$$\mathbb{P}(|(\bar{\mathbf{f}}_\varepsilon^{(\ell)})_{k_j}| < 1 \,|\, |(\mathbf{f}_\varepsilon)_\ell| \geq 1),$$

for $j = 1, \ldots, 6$, and found that the probability that all the six lines identified by the large deviations approach actually fail in the pre-limit is equal to

$$\begin{aligned}&\mathbb{P}(|(\bar{\mathbf{f}}_\varepsilon^{(\ell)})_{k_j}| \geq 1 \,\forall j = 1, \ldots, 6 \,|\, |(\mathbf{f}_\varepsilon)_\ell| \geq 1)\\ &\geq 1 - \sum_{j=1}^{6} \mathbb{P}(|(\bar{\mathbf{f}}_\varepsilon^{(\ell)})_{k_j}| < 1 \,|\, |(\mathbf{f}_\varepsilon)_\ell| \geq 1) = 0.9987.\end{aligned}$$

4.B.3 German network: System security vs. system cost

In order to model a heavily-loaded but not overloaded system, in the OPF we scale the true line limits $\bar{f}_\ell$ by a contingency factor of $\lambda \in (0, 1)$. This is a common practice in power engineering that allows room for reactive power flows and stability reserve.

We explore the trade-off between system security and system cost, by varying the contingency factor λ in the range $\lambda \in [0.7, 1]$. We evaluate system security by means of the large deviations approximation for the failure probability of a given line ℓ,

$$\mathrm{pr}^{(\lambda)}(\ell) = \exp(-I_\ell(\lambda)), \tag{4.24}$$

where we emphasize the dependency on λ, and we use the average Locational Marginal Price (LMP, Section 1.3.2) and the maximum LMP at the grid nodes as metrics of system costs.

Fig. 4.9 reports the results corresponding to the same setting as in Fig. 4.7. From this graph one can, for instance, immediately infer that making line 27 ten times as safe will roughly cost 1 €/MWh on average, while the increase in cost in terms of maximum price can be much more significant. This example shows how our large deviations theoretical framework can be a valuable tool to help designing a safe and reliable network at minimal cost.

Reducing the security margin does not only influence the average LMPs and system costs, but also their geographical distribution. Fig. 4.10 shows the LMPs

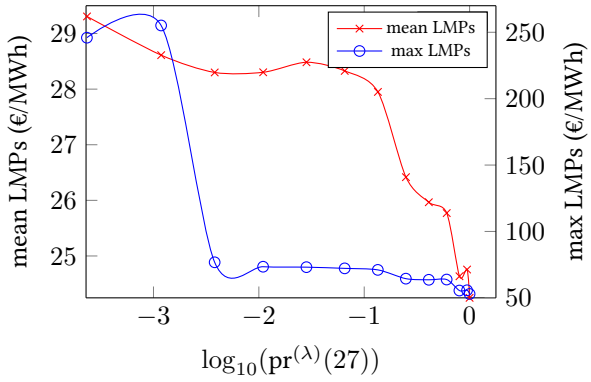


Figure 4.9: Average LMP (scale on left) and Maximum LMP (right) vs. $\log_{10}(\mathrm{pr}^{(\lambda)}(27)) = \log_{10}(\exp(-I_{27}))$, for the German network at 4pm.

for two values of λ, one corresponding to a low effective limit/large security margin system ($\lambda = 0.7$) and the other to a large effective limit/low security margin system ($\lambda = 0.95$). We can see how a more conservative system results in higher LMPs, especially in the south and south-west part of Germany, while in northern Germany the difference is less pronounced. Quoting [26], this phenomenon can be explained by the fact that "transmission bottlenecks in the middle of Germany prevent the transportation of this cheap electricity to the South, where more expensive conventional generators set the price".

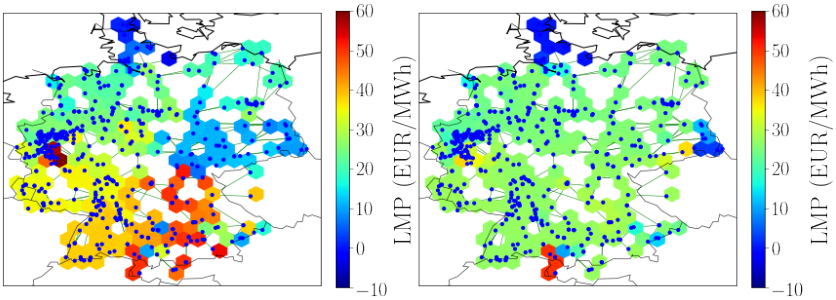


Figure 4.10: Geographical distribution of LMPs for $\lambda = 0.7$ (left) and $\lambda = 0.95$ (right), at 4pm.

Furthermore, Fig. 4.11 shows that reducing the system security margin does not only increase the likelihood of an overload, but it also increases the number of lines with a large enough overload probability.

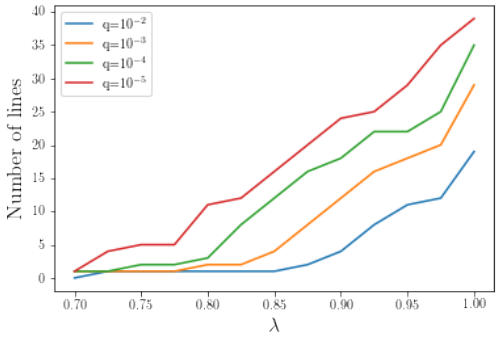


Figure 4.11: Number of lines ℓ with overload probability $\text{pr}^{(\lambda)}(\ell) \geq q$.

4.B.4 IEEE test cases: analysis of classical vs. emergent failures

As illustrated earlier, the most likely power injections configuration leading to the emergent failure of a given line can be used in combination with the power flow redistribution rules to generate the failures triggered by that initial scenario. By repeating this procedure for all lines, one can obtain insightful statistics of the first two stages of emergent cascading failures (ec) and compare them with those of classical cascading failures (cc), obtained using nominal power injection values rather than the most likely ones and deterministic removal of the initial failing line. We perform numerical experiments using IEEE test grids. Since several IEEE test-cases do not report realistic transmission limits, line capacities are taken to be proportional to the average absolute power flow on the corresponding lines, i.e., $\bar{f}_\ell = (1+\alpha)|\nu_\ell|$, where ν_ℓ is a nominal value provided in the dataset, $\alpha = 0.25$ and $\mathbf{\Sigma}_p$ is the identity matrix.

Graph	% joint failures	$\mathbb{E}(F_1^{\text{ec}})$	$\mathbb{E}(F_2^{\text{ec}})$	$\mathbb{E}(F_2^{\text{cc}})$
IEEE 14	65.0%	4.40	8.40	4.95
IEEE 30	97.6%	3.73	9.88	4.95
IEEE 39	80.4%	4.78	11.39	4.85
IEEE 57	88.5%	8.00	19.00	10.44
IEEE 96	72.2%	6.70	21.47	7.31
IEEE 118	91.6%	10.40	24.53	7.56
IEEE 300	87.0%	18.13	39.19	7.42

Table 4.4: Percentage of joint failures in emergent cascades and average number of failed lines F_1 up to stage 1 and F_2 up to stage 2 for emergent cascades (ec) and classical cascades (cc) for some IEEE test systems.

As shown in Table 4.4, emergent cascades have a very high percentage

of joint failures and an average number of failures in the first cascade stage much larger than one (in classical cascades only one line is removed in the first cascade stage). Furthermore, the expected total number of failed lines up to the second cascade stage is significantly larger for emergent cascades than for classical cascades. Lastly, failures propagate in emergent cascades on average a bit less far than in classical cascades, as illustrated by the statistics of the failure jumping distance in Table 4.5.

Graph	$\mathbb{E}(D^{\text{ec}})$	$\mathbb{E}(D^{\text{cc}})$	$c_v(D^{\text{ec}})$	$c_v(D^{\text{cc}})$
IEEE 14	0.388	0.987	0.600	1.050
IEEE 30	0.754	1.198	0.879	1.115
IEEE 39	0.898	1.633	0.891	1.149
IEEE 57	1.210	2.507	0.863	1.415
IEEE 96	1.450	1.781	0.879	0.946
IEEE 118	0.679	1.638	0.745	1.169
IEEE 300	1.408	2.580	0.806	1.081

Table 4.5: Average and coefficient of variation of the failure jumping distance D in stage 2 both for emergent cascades (ec) and classical cascades (cc). The distance between two lines is measured as the shortest path between any of their endpoints.

Our approach also gives a constructive way to build the so-called "influence graph" [87, 88, 152], in which a directed edge connects lines ℓ and ℓ' if the failure of the line ℓ triggers (simultaneously or after redistribution) that of line ℓ'. Fig. 4.12 shows an example of influence graph built using our large deviations approach. The *cliques* of the influence graph (i.e., its maximal fully connected subgraphs) can then be used to identify clusters of cosusceptable lines [207], which are the lines that statistically fail often in the same cascade event.

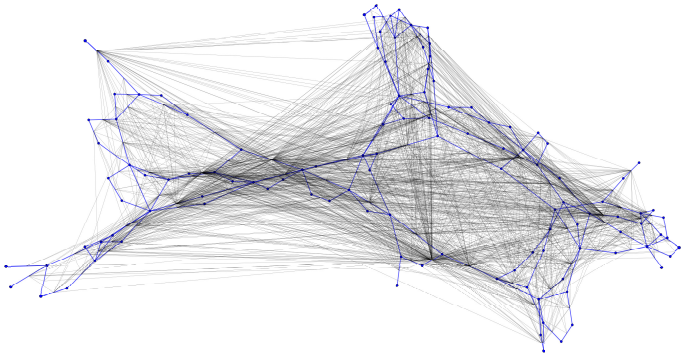


Figure 4.12: The influence graph of the IEEE 118-bus test system (in black) built using the first two stages of all cascade realizations has a different structure than the original network (in blue).

Chapter 5

Emergence of Scale-Free Blackout Sizes

Contents

In this chapter, we focus on an important macroscopic feature of blackouts in power grids, namely that the total size of a blackout is scale-free, i.e., its probability distribution has a Pareto law, and propose a novel explanation for this phenomenon.

We model power grids as graphs with *heavy-tailed* sinks, which represent demand from cities, and study cascading failures on such graphs. Our analysis links the scale-free nature of blackout size to the scale-free nature of city sizes, contrasting previous studies suggesting self-organized criticality (Section 1.4.2) as a possible explanation. In particular, we show that the tail behavior of the blackout size distribution is completely determined by the city size distribution.

Our results are based on a new mathematical framework that combines a structural model of the power grid with rare event analysis for heavy-tailed distributions, and are validated on the German transmission grid, as well as various synthetic networks.

An implication of these results is that, as we elaborate more extensively in Section 5.5, the classical approach of enhancing power grids' resilience (e.g. by investing in network upgrades [57, 206]) would only lead to a modest decrease in the likelihood of big blackouts. Conversely, our insights suggest that it would be more effective to strengthen the resilience of cities by investing in responsive measure, such as energy storage, that enable consumers to mitigate the consequences of large blackouts.

Finally, our insights are not limited to power grids, and offer new ways of approaching scale-free phenomena in other transportation networks as well.

Chapter outline: The rest of the chapter is organized as follows. In Section 5.1 we describe our model for the power grid, formulate our research question, and present results from the statistical analysis of historical blackouts. In Section 5.2 we present the mathematical framework that we use to model the cascading failure process, and state a property of Pareto distributions that is a crucial ingredient of our analysis. Next, we provide the rigorous mathematical derivation of our results in Section 5.3, and present the simulation experiments in Section 5.4. Finally, we conclude in Section 5.5. Details on the statistical analysis and modeling, as well as extended proofs, are reported in Appendices 5.A, 5.B and 5.C.

5.1 System model and problem formulation

Following the model description of Section 1.2.2, we view the power grid as a connected graph $\mathcal{G} = \mathcal{G}(\mathcal{N}, \mathcal{E})$ with $|\mathcal{N}| = n, |\mathcal{E}| = m$, where nodes $\mathcal{N}$ represent buses, which are connected by edges $\mathcal{E}$ modeling transmission lines. Moreover, we let $\mathbf{g}, \mathbf{d} \in \mathbb{R}^n$ and $\mathbf{p} = \mathbf{g} - \mathbf{d} \in \mathbb{R}^n$ represent the nodal (active) generation, demand, and net power injection vectors, respectively. We assume that every node $i \in \mathcal{N}$ represents a *city* with *size* X_i, where we define the size of a city as the number of inhabitants, and we denote by $\mathbf{X} = (X_1, \dots, X_n) \in \mathbb{R}^n$ the vector of city sizes.

We consider a static framework where each inhabitant demands one unit of energy, therefore identifying the energy demand of a city with its size, i.e.

$$d_i = X_i \text{ for every } i = 1, ..., n. \tag{5.1}$$

In view of Eq. (5.1), we will use the notations $\mathbf{X}$ and $\mathbf{d}$ interchangeably. Under the DC approximation, the power flows $\mathbf{f} \in \mathbb{R}^m$ are given by $\mathbf{f} = \widehat{\mathbf{V}}\mathbf{p}$, where the PTDF matrix $\widehat{\mathbf{V}}$ was defined in Eq. (1.19). In this context, the DC power flow model has been shown to be accurate in describing the evolution of the cascade that led to the 2011 San Diego blackout [13]. Some useful properties of matrix $\widehat{\mathbf{V}}$, that will be used later, are reported in Appendix 5.B.

Under normal conditions, the power grid is a single fully functioning network with balanced supply and demand, i.e., $\mathbf{e}^\top(\mathbf{g} - \mathbf{d}) = 0$, where $\mathbf{e} = (1, \dots, 1)^\top \in \mathbb{R}^n$. Following [14], a typical way for a large blackout to occur is as follows: after several line failures, the network breaks into disconnected subnetworks, referred to as *islands* in the rest of the chapter. In particular, suppose that after several line failures the network disconnects in two islands, $\mathcal{I}_1$ and $\mathcal{I}_2 = \mathcal{N} \setminus \mathcal{I}_1$. In general, the balance between supply and demand is not guaranteed to hold in the individual islands, i.e. $\sum_{i\in\mathcal{I}_1}(g_i - d_i) = -\sum_{i\in\mathcal{I}_2}(g_i - d_i) \neq 0$. Without loss of generality, assume that $\sum_{i\in\mathcal{I}_1}(g_i - d_i) < 0$, so that $\mathcal{I}_1$ faces a power shortage. We define the size of the blackout, *up to the first disconnection*, in terms of the amount of load that needs to be shed in $\mathcal{I}_1$ in order to restore balance, i.e.

$$S = \left| \sum_{i\in\mathcal{I}_1} (g_i - d_i) \right| = \sum_{i\in\mathcal{I}_1} (d_i - g_i) > 0, \tag{5.2}$$

which is equivalent to the *number of customers affected*, in view of Eq. (5.1). After this first disconnection, the cascade may progress leading to a network having several disconnected islands. The *total blackout size* is defined as the cumulative load shed in each island facing a shortage until the cascading failure process ends or, equivalently, as the total number of customers affected by the blackout in view of Eq. (5.1). A rigorous description of our model for blackouts is given in Section 5.2.

Research question: It is well-documented [38, 57, 86, 39] that blackout sizes, expressed in terms of number of customers affected, are scale-free, i.e., their distribution is consistent with that of a Pareto random variable. A Pareto-distributed random variable X with minimum value $x_{\min} > 0$ and tail exponent $\alpha > 0$ is described by its complementary cumulative distribution function (CCDF)

$$\bar{F}(x) = \mathbb{P}(X > x) = \left(\frac{x}{x_{\min}}\right)^{-\alpha}, \; x \geq x_{\min}. \tag{5.3}$$

The expected value of X is equal to $(\alpha x_{\min})/(\alpha - 1)$ if $\alpha > 1$, and ∞ otherwise. Whenever we are only interested in the tail index α, we will use the notation

$\mathbb{P}(X > x) \approx Cx^{-\alpha}$, where $\approx$ means that the ratio of both quantities approaches 1 as $x \to \infty$, and $C > 0$ is a constant. The results of the statistical analysis for historical blackouts in the US are reported in Table 5.1. In particular, we observe that the estimated tail index is greater than 1, implying a finite mean. The statistical analysis is based on the PLFIT method in [42], and details are reported in Appendix 5.A.

Dataset	N	n_{tail}	$\hat{\alpha}(\hat{x}_{\text{min}})$	$\hat{x}_{\text{min}}$	KS p-value
US city sizes ($\times 10^3$)	19447	580	1.37 ± 0.08	52.5 ± 11.6	0.76
US blackout sizes ($\times 10^3$)	1341	448	1.31 ± 0.08	140 ± 31.3	0.32

Table 5.1: PLFIT statistics for US city [42] (based on the 2000 Census) and blackout sizes [184]. n_{tail} is the number of data points x_i greater or equal than $\hat{x}_{\text{min}}$. Standard deviations obtained via nonparametric bootstrap with 1000 repetitions, and KS p-value denotes the result of the Kolmogorov-Smirnov goodness-of-fit test (Section 5.A.1).

The central thesis of this chapter is that the scale-free nature of city sizes in the network is what drives the scale-free nature of blackout sizes. In particular, we will show that city sizes X and blackout sizes S both have Pareto distributions with similar tail behavior, i.e.

$$\mathbb{P}(S > x) \approx C_1 x^{-\alpha}, \mathbb{P}(X > x) \approx C_2 x^{-\alpha}, \tag{5.4}$$

for $C_1, C_2 > 0$. For the case of the US blackout sizes and city sizes, we corroborate this with historical data as summarized in Fig. 5.1 and Table 5.1, which shows that the distribution of city sizes is also scale-free [42], and that the tail parameters α for blackout and city sizes distributions are remarkably similar (blackout sizes: $\alpha = 1.31 \pm 0.08$; city sizes: $\alpha = 1.37 \pm 0.06$). We refer to Appendix 5.A for details on the historical data analysis and the statistical procedure.

5.2 Cascading failures model

In what follows, we formally state our results by introducing a new mathematical framework that captures the salient characteristics of actual power system dynamics [14] and sheds light on the connection between blackout and city sizes.

In accordance with data analysis (see Table 5.1), we model the size of a city with a Pareto distribution with tail index α, i.e.

$$\mathbb{P}(X > x) \approx C_2 x^{-\alpha}, \tag{5.5}$$

for a constant $C_2 > 0$, and we assume that city sizes are independent and identically distributed. For convenience, we relabel the nodes such that X_1 represents the largest city.

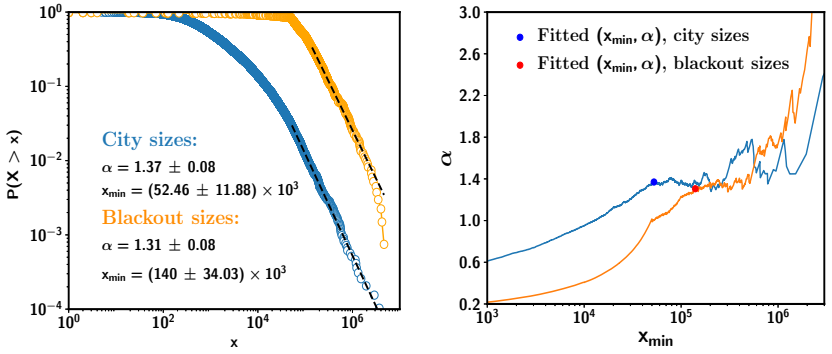


Figure 5.1: a) Pareto tail behavior of US city [42] and blackout sizes [184] in the region $x \geq x_{\min}$. Estimates for α and $x_{\min}$, along with standard deviations, are based on PLFIT [42], and are statistically significant based on the Kolmogorov-Smirnov goodness-of-fit test (see Appendix 5.A.1). Points depict the empirical complementary cumulative distribution function (CCDF); Solid line depicts the CCDF of a Pareto distribution with parameters $\alpha, x_{\min}$. b) Visualization of the estimated α obtained by only considering values in the region $x \geq x_{\min}$, as a function of $x_{\min}$. The PLFIT-estimated $x_{\min}$ for city sizes (blue dot) and blackout sizes (red dot) lie within a region where the values of α are relatively stable, corroborating the results of the PLFIT procedure.

For the operations of our network, we employ the Direct Current Optimal Power Flow (DC-OPF) formulation, which plays a central role in the daily operations of actual power systems [14, 33]. As described in Section 1.3.1, the OPF is a cost-minimization procedure that determines the optimal generation schedule satisfying demand/supply balance and network operating constraints, such as generator and transmission lines limits, under the assumptions of the DC approximation.

5.2.1 Planning, operational and emergency problems.

In order to obtain a fundamental, causal understanding of the correlation between blackout sizes and city sizes using the DC approximation model, we require a framework that adequately sets the transmission lines limits by taking into account the topology of the network and the distribution of city sizes. In addition, we need to specify a mechanism that causes the initial line failure, as well as which lines possibly fail next after the power flow redistribution. For this purpose, we consider a framework that consists of three problems: the *planning* problem, the *operational* problem, and the *emergency* problem. Next, we explain our framework in more detail, followed by listing some vital properties.

The planning problem. The planning problem refers to the design of the power network, and in particular, how the line limits $\bar{\mathbf{f}}$ for the next phase, the operational problem, are set.

Recall that, in our static framework, each inhabitant demands one unit of energy, i.e. $d_i = X_i$ for every $i = 1, ..., n$, and that city sizes X_i's are i.i.d. and have a Pareto tail, as given by Eq. 5.5. We assume that the generation cost function at node i is a strictly convex and quadratic function of g_i, and that generator limits do not pose an effective constraint in the DC-OPF. Specifically, we assume that $J_i(g_i) = g_i^2/2$, $i = 1, ..., n$, and $\underline{g}_i = -\infty, \bar{g}_i = \infty$, $i = 1, ..., n$. Moreover, we assume that line limits do not pose an effective constraint. [1]

These assumptions ensure that the generation is spread among the cities as equally as possible, and that the resulting line power flows, as well as the operational line limits (defined later in Eq. (5.6)), are also Pareto-tailed with the same exponent as the city sizes.

The solution of the DC-OPF in the absence of any generator and transmission line limits is

$$\mathbf{g}^{(\text{planning})} = \Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big)\mathbf{e}.$$

The associated flow vector is given by

$$\mathbf{f}^{(\text{planning})} = \widehat{\mathbf{V}}(\mathbf{g}^{(\text{planning})} - \mathbf{X}) = -\widehat{\mathbf{V}}\mathbf{X},$$

where we used that $\widehat{\mathbf{V}}\mathbf{g}^* = \mathbf{0}$ (Lemma 5.4). Next, in the operational problem, we set operational line limits as a fraction λ of the planning problem power flows

$$\bar{f}_\ell = \lambda|\mathbf{f}^{(\text{planning})}| = \lambda|(\widehat{\mathbf{V}}\mathbf{X})_\ell|, \qquad \ell = 1, ..., m, \tag{5.6}$$

where $\lambda \in (0, 1]$ is a safety tuning parameter.

The operational problem. The operational problem consists in solving the DC-OPF with respect to the operational line limits Eq. (5.6) to obtain the generation vector $\mathbf{g}$. That is, we solve

$$\min_{g\in\mathbb{R}^n} \sum_{i=1}^{n} g_i^2/2 \tag{5.7}$$

$$\text{s.t.} \quad \sum_{i=1}^{n} g_i = \sum_{i=1}^{n} X_i, \tag{5.8}$$

$$-\bar{\mathbf{f}} \le \widehat{\mathbf{V}}(\mathbf{g} - \mathbf{X}) \le \bar{\mathbf{f}} \tag{5.9}$$

[1] We remark that the design phase's purpose is to set line limits for the operational phase.

Observe that the line limit constraints $-\bar{\mathbf{f}} \leq \widehat{\mathbf{V}}(\mathbf{g} - \mathbf{X}) \leq \bar{\mathbf{f}}$ can be rewritten, in view of Eq. (5.6), as

$$\widehat{\mathbf{V}}\mathbf{X} - \lambda \left|\widehat{\mathbf{V}}\mathbf{X}\right| \leq \widehat{\mathbf{V}}\mathbf{g} \leq \widehat{\mathbf{V}}\mathbf{X} + \lambda \left|\widehat{\mathbf{V}}\mathbf{X}\right| \tag{5.10}$$

where $|\widehat{\mathbf{V}}\mathbf{X}|$ denotes the vector with elements $(|\widehat{\mathbf{V}}\mathbf{X}|)_j = |(\widehat{\mathbf{V}}\mathbf{X})_j|$, $j = 1, ..., m$.

The emergency problem. Finally, the emergency problem concerns how an initial disturbance propagates through the network. We assume that the initial disturbance is caused by a single line failure, chosen uniformly at random over all lines. We point out that our framework can be extended to multiple initial line failures, or adapted to deal with generator failures. The initial line failure changes the topology of the grid and causes a global redistribution of network flows according to power flow physics (see Appendix 5.B).

A subsequent line failure occurs whenever there is at least one line such that its *emergency line limit* is exceeded. The emergency line limit is defined as

$$F_\ell = \lambda^* \bar{f}_\ell, \tag{5.11}$$

with $\bar{f}_\ell$ being the (conservatively chosen) operational line limits, and $\lambda^* > 1$ is a constant. A canonical choice is $\lambda^* = 1/\lambda$.

We assume that line failures occur subsequently, and occur first at the line where the relative exceedance is largest. Whenever line failures cause the network to disconnect in multiple islands, we assume that the energy balance is restored by proportionally lowering either generation or demand at all nodes, as described in more details in Section 5.3.2.

This change in power injections alters the power flows as well, and the process can continue resulting in a cascading effect. More specifically, before the initial disturbance occurs, the network flows are given by $\widehat{\mathbf{V}}(\mathbf{g} - \mathbf{d})$, where $\mathbf{g}$ is the solution of the DC-OPF in the operational problem, and $\mathbf{d} = \mathbf{X}$. After any line failure, we check whether this causes the network to disconnect, and if so, we proportionally lower the generation in one component and the demand in the other component such that demand and generation are balanced in the two disconnected components. The network flows are updated according to the laws of physics in every component. That is, the removal of one or more lines yields a modified matrix $\widetilde{\widehat{\mathbf{V}}}$ (see Appendix 5.B), and possibly modified generation $\tilde{\mathbf{g}}$ and demand $\tilde{\mathbf{d}}$. The line flows are given by $\widetilde{\widehat{\mathbf{V}}}(\tilde{\mathbf{g}} - \tilde{\mathbf{d}})$. This cascading failure process continues until the emergency line limits F_ℓ of all surviving lines are sufficient to carry the power flows.

This iterative process may lead to the disconnection of the network in multiple disconnected sets, or islands. Within each island, we alter the generation and demand to restore the power balance. Observe that there is at least one island that experiences a power shortage and needs to shed load. Our object of interest is the total amount of load that is shed during the process.

In particular, define $\mathcal{A}_1 \subseteq \mathcal{N}$ to be the island that contains the city with the largest demand after the cascade has taken place. We point out that the set $\mathcal{A}_1$ is random, and in particular, $\mathcal{A}_1 = \{1, ..., n\}$ if the cascade stops without causing network disconnections.

The mismatch between generation and demand in the component $\mathcal{A}_1$, which contains the city with highest power demand, is given by

$$S = \left| \sum_{i \in \mathcal{A}_1} (X_i - g_i) \right| = \sum_{i \in \mathcal{A}_1} (X_i - g_i) > 0. \tag{5.12}$$

In general, the total size of the blackout can be greater than S, as there may be additional islands experiencing power shortages. Thanks to properties of the Pareto distribution, however, it turns out that S is a good approximation for the total size of the blackout, as it yields exactly the same limiting behavior. We study this notion in more details in the next sections.

5.2.2 Principle of a single city with large demand

A key property in our framework is that the tail of the blackout distribution is dominated by the scenario where there is a single city that has a large power demand, while the demand of the other cities is negligible. To formalize this notion, write $d_1 = \max\{d_1, ..., d_n\}$ with $d_i, i = 1, ..., n$, independent and identically Pareto distributed power demands. Note that for every $\epsilon > 0$,

$$\mathbb{P}(S > x) = \mathbb{P}\left(S > x; \sum_{i=2}^{n} d_i < \epsilon d_1\right) + \mathbb{P}\left(S > x; \sum_{i=2}^{n} d_i \geq \epsilon d_1\right).$$

The next lemma shows that the second term in the right-hand side is negligible.

Lemma 5.1. *Suppose d_i, $i = 1, ..., n$ are independent and identically Pareto distributed with tail exponent $\alpha > 0$, and write $d_1 = \max\{d_1, ..., d_n\}$. For every $\epsilon > 0$, as $x \to \infty$,*

$$\mathbb{P}\left(S > x; \sum_{i=2}^{n} d_i \geq \epsilon d_1\right) = O\left(x^{-2\alpha}\right). \tag{5.13}$$

In other words, Lemma 5.1 implies that if for some $\epsilon > 0$ sufficiently small, the approximation

$$\mathbb{P}\left(S > x; \sum_{i=2}^{n} d_i < \epsilon d_1\right) \approx Cx^{-\alpha}$$

holds for some constant $C \in (0, \infty)$, then the only likely way to have a large blackout is when there is a single city that has a large demand.

As a result of Lemma 5.1, in order to understand how large blackouts occur, it suffices to grasp the behavior in the special case where there is a single city with a large demand d_1, while all other demands are negligible, i.e., $\sum_{i=2}^{n} d_i < \epsilon d_1$.

To this end, in the next section we first focus on the very special case where $\mathbf{d} = y\mathbf{e}_1$, i.e., $d_1 = y > 0$ and $d_j = 0$ for $j > 1$, for which the blackout size S has a particularly simple expression. In view of Lemma 5.1, this special case describes some form of limiting behavior for the general case, which will be analyzed extensively in Section 5.3.

5.2.3 Blackout size in the case $\mathbf{d} = y\mathbf{e}_1$.

In this section, we show that in the special case $\mathbf{d} = y\mathbf{e}_1$, $y > 0$, the operational problem and has a closed-form solution, given by

$$g_1 = y\Big(1 - \lambda\frac{n-1}{n}\Big),\ g_j(\lambda) = \lambda\,\frac{y}{n} \quad \text{for } j \geq 2. \tag{5.14}$$

In turn, this allows us to get a simple expression for the blackout size S.

First, we consider the planning problem. In the absence of any generator and transmission line limits, the solution of the planning OPF is $\mathbf{g}^* = \frac{y}{n}\mathbf{e}$, with associated flow vector $\mathbf{f}^* = \widehat{\mathbf{V}}(\mathbf{g}^* - y\mathbf{e}_1) = -y\widehat{\mathbf{V}}\mathbf{e}_1$, where we used that $\widehat{\mathbf{V}}\mathbf{g}^* = \mathbf{0}$ (see Lemma 5.4).

Therefore, the operational problem reduces to

$$\min_{g\in\mathbb{R}^n} \quad \sum_{i=1}^{n} g_i^2/2 \tag{5.15}$$

$$\text{s.t.} \quad \mathbf{e}^\top\mathbf{g} = \mathbf{e}^\top\mathbf{e}_1 = y, \tag{5.16}$$

$$-\lambda y|\widehat{\mathbf{V}}\mathbf{e}_1| \leq \widehat{\mathbf{V}}(\mathbf{g} - \mathbf{d}) \leq \lambda y|\widehat{\mathbf{V}}\mathbf{e}_1|, \tag{5.17}$$

which we will denote by $P(\lambda, y)$. Lemma 5.2 shows that the solution of $P(\lambda, y)$ is of closed form.

Lemma 5.2. *Let $\lambda \in (0,1)$. Let $\mathcal{G}$ be assigned the orientation such that $\widehat{\mathbf{V}}\mathbf{e}_1 \geq \mathbf{0}$. Then, the solution of $P(\lambda, y)$ is given by*

$$\mathbf{g}(\lambda) = \lambda \frac{y}{n}\mathbf{e} + y(1-\lambda)\mathbf{e}_1,$$

i.e., $g_1(\lambda) = y(1-\lambda\frac{n-1}{n})$ and $g_i(\lambda) = \lambda\frac{y}{n}$ for all $i = 2, \ldots, n$. The corresponding line flows are at capacity and are given by $\mathbf{f}(\lambda) = -y\lambda\widehat{\mathbf{V}}\mathbf{e}_1$.

The proof is reported in Appendix 5.C.

Finally, we solve the emergency problem, which leads to a simple expression for the blackout size. Note that whenever there is a network disconnection, the component that does not contain node 1 has zero power demand, and hence the generation at every node in that component must be lowered to zero in order to restore balance. Evidently, no consecutive failures occur in this component, implying that the total amount of load that is shed equals the power imbalance in just the component containing the biggest city, $\mathcal{A}_1$, and reads

$$S = \sum_{i \in \mathcal{A}_1} (X_i - g_i) = \sum_{j \notin \mathcal{A}_1} (g_j - X_j) = \sum_{i \notin \mathcal{A}_1} \frac{\lambda}{n} = \lambda \frac{n - |\mathcal{A}_1|}{n} y. \tag{5.18}$$

Indeed, the demand at node 1 is reduced by the number of nodes that disconnect from this component times $y\lambda/n$, since $g_j = y\lambda/n$ for every $j \geq 2$. Next, we will see how this special case can be used to derive the distribution of the tail of blackout sizes in the general case.

5.3 Blackout size in the general case

5.3.1 Intuitive argument

Before proceeding with the rigorous mathematical analysis, it is helpful to get an intuitive understanding of the main idea underlying our results. For this reason, we now provide a sketch of our argument, referring to Sections 5.3.2 and 5.3.3 for the full derivation. First, observe that in view of Lemma 5.1, the case where $\sum_{i=2}^{n} d_i \leq \varepsilon d_1$ (for arbitrarily small $\varepsilon > 0$) is, in certain settings, the only likely way to have a large blackout.

Therefore, in the following derivation, we will consider the limiting regime $\mathbf{d} \to y\mathbf{e}_1$, where we recall that $d_1 = \max\{d_1, \ldots, d_n\}$ is the city with largest demand. By conditioning on the size of the island $\mathcal{A}_1$ containing the biggest city at the end of the cascade, we have

$$\begin{aligned}\mathbb{P}(S > x) &= \sum_{j=1}^{n} \mathbb{P}\left(S > x \,\middle|\, |\mathcal{A}_1| = j\right) \mathbb{P}\left(|\mathcal{A}_1| = j\right) \\ &\approx \sum_{j=1}^{n-1} \mathbb{P}\left(\lambda \frac{n-j}{n} d_1 > x \,\middle|\, |\mathcal{A}_1| = j\right) \mathbb{P}\left(|\mathcal{A}_1| = j\right) \\ &\approx C_1 x^{-\alpha},\end{aligned}$$

where

$$C_1 = nK \sum_{j=1}^{n-1} \mathbb{P}\left(|\mathcal{A}_1| = j\right) \lambda^{\alpha} (1 - j/n)^{\alpha} \in [0, \infty). \tag{5.19}$$

Intuitively, the second approximation in the above derivation follows by considering the limiting regime $\mathbf{d} \to y\mathbf{e}_1$, recalling that in the case $\mathbf{d} = y\mathbf{e}_1$ the blackout size is equal to $\lambda \frac{n-|\mathcal{A}_1|}{n} y$, and invoking an important continuity property described in Lemma 5.1. Finally, in the last step we used the property

$$\mathbb{P}(\max\{X_1, ..., X_n\} > x) \approx n\mathbb{P}(X > x) \approx nKx^{-\alpha},$$

that holds for Pareto tails [156]. Therefore, if there is a strictly positive probability $\mathbb{P}(|\mathcal{A}_1| \leq n-1) = \sum_{j=1}^{n-1} \mathbb{P}(|\mathcal{A}_1| = j) > 0$ that a cascade creates additional islands, then $C_1 > 0$ and the total amount of load shed has a Pareto tail with the same exponent as the city size distribution Eq. (5.4). The next subsections are devoted to the rigorous derivation of this result.

5.3.2 Convergence of cascade sequence and continuity properties

First, note that without loss of generality, we can always normalize our framework by dividing all parameters (generation, line limits, etc.) by the sum of all power demands. This yields an equivalent setting where the total power demand equals one, i.e., $\mathbf{d} = \mathbf{e}_1$. Therefore, we will normalize $y = 1$ in the derivation below. In the rest of this section, we first rigorously define our framework for the cascade evolution, and then we show how, under some assumptions, for all demand vectors $\mathbf{d}$ for which $\mathbf{d} \to \mathbf{e}_1$ as $\epsilon \downarrow 0$, the sequence of subsequent line failures (i.e., the cascade evolution) converges to the sequence of line failures as

if the demand vector would have been $\mathbf{d} = \mathbf{e}_1$. The operational DC-OPF

$$\min_{\mathbf{g}\in\mathbb{R}^n} \frac{1}{2}\mathbf{g}^\top \mathbf{g} \tag{5.20}$$

$$\text{s.t.} \quad \mathbf{e}^\top \mathbf{g} = \mathbf{e}^\top \mathbf{d}, \tag{5.21}$$

$$|\widehat{\mathbf{V}}(\mathbf{g}-\mathbf{d})| \le \lambda \left|\widehat{\mathbf{V}}\mathbf{d}\right|, \tag{5.22}$$

is a strictly convex optimization problem, and since $\mathbf{g} = \lambda\bar{\mathbf{d}}\mathbf{e} + (1-\lambda)\mathbf{d}$ is a feasible point, the feasible set of this optimization problem is nonempty. Therefore, for each demand vector $\mathbf{d}$, there exists a unique optimal solution $\mathbf{g}^*(\mathbf{d})$.

If we view $\mathbf{d}$ as a *parameter* of the problem, then the optimization problem defined by Eqs. (5.20)-(5.22) is an instance of a *multi-parametric quadratic programming* (mp-QP) problem with a strictly convex objective function, for which it is known that the optimal solution $\mathbf{g}^*(\mathbf{d})$ is a continuous function of the parameter vector $\mathbf{d}$ (Theorem 1, [181]). This continuity property will be used extensively in the rest of the section.

We assume in our framework that line failures occur subsequently, i.e., a next line failure occurs at the line where the line limit is relatively most exceeded. Recall that F_j denotes the *emergency* line limit of line $j \in \mathcal{E}$, and is given by (taking $\lambda^* = \frac{1}{\lambda}$)

$$F_j = \lambda^*\lambda|(\widehat{\mathbf{V}}\mathbf{d})_j| = |(\widehat{\mathbf{V}}\mathbf{d})_j|.$$

We write $f_j^{(m)}$ as the flow on line j after the failure of the first $m-1$ lines and after the load/generation shedding took place, where we use the convention that $f_j^{(1)}$ denotes the flow on line j when no initial disturbance has occurred yet, and $f_j^{(m)} = 0$ if line j has already failed before the m-th step of the cascading failure process. The cascade is initiated by the random failure of line $\ell = \ell^{(1)}$. For $m \ge 2$, the m-th line to fail is given by

$$\ell^{(m)} = \arg\max_{j\in\mathcal{A}^{(m)}}\left\{\frac{|f_j^{(m)}| - F_j}{F_j}\right\} = \arg\max_{j\in\mathcal{A}^{(m)}}\left\{\frac{|f_j^{(m)}|}{F_j}\right\}, \tag{5.23}$$

where $\mathcal{A}^{(m)} = \{j : |f_j^{(m)}| \ge F_j\}$ is the set of lines that exceed the limit.

Remark 5.1. *Note that the line limits and line flows depend on* $\mathbf{d}$ *and* λ *through the operational OPF, so that the sequence of subsequent failure depends on* $\mathbf{d}, \lambda$*, and on the initial failure* $\ell = \ell^{(1)}$*. That is,*

$$F_j = F_j(\mathbf{d}),\ f_j^{(m)} = f_j^{(m)}(\mathbf{d},\lambda),\ \mathcal{A}^{(m)} = \mathcal{A}^{(m)}(\ell,\lambda,\mathbf{d}),\ \ell^{(m)} = \ell^{(m)}(\ell,\lambda,\mathbf{d}).$$

For notational compactness, we do not write the dependency on $\mathbf{d}$, λ *and* ℓ.

We remark that if $|\mathcal{A}^{(m)}| = 0$, no more line failures occur. Technically, it is also possible that $|\mathcal{A}^{(m)}| > 1$ and, hence, the subsequent line failure next needs to be chosen from a set of multiple lines. We exclude the cases that do not yield unique maximizers from our framework.

Assumption 5.1. *For all lines* j*, the ratios between redistributed flows and line limits*

$$\frac{|f_j^{(m)}(\mathbf{e}_1)|}{F_j(\mathbf{e}_1)} = \frac{\lambda|(\widehat{\mathbf{V}}^{(m)}\mathbf{e}_1)_j|}{|(\widehat{\mathbf{V}}\mathbf{e}_1)_j|}$$

are all different for all $m \geq 2$*, where* $\widehat{\mathbf{V}}^{(m)}$ *denotes the PTDF matrix for the remaining network after* $m-1$ *failures have taken place.*

Assumption 5.1 is needed to ensure the uniqueness of the maximizer in Eq. (5.23), and guarantees that whenever the first line failure ℓ and the parameter λ are known, the cascade sequence for the demand vector $d = \mathbf{e}_1$ is unique and deterministic. This assumption is added for technical convenience, and we stress that our results hold more generally. In particular, this assumption rules out certain network topologies with some form of symmetry, but we can slightly adapt the framework to deal with these cases as well.

That is, suppose that $|\mathcal{A}^{(m)}| > 1$ for some $m \in \mathbb{N}$ and the set $\mathcal{A}^{(m)}$ consists only of lines that are indistinguishable from one another (lines that are 'symmetric'). Since nodal demands are independent and identically distributed, this implies that each of these lines has an equal probability of being the line that fails next. By the symmetry of the network topology, regardless of which line is chosen to fail next, the resulting networks after the cascade are indistinguishable.

Let $\mathcal{C} = \{\ell^{(1)}, \dots, \ell^{(T)}\}$ be a cascade sequence, where $\ell^{(T)}$ is the last failure before the cascade stops. In view of Assumption 5.1, such a sequence is uniquely determined by the first failure $\ell^{(1)}$ and by the demand vector $\mathbf{d}$ and by λ, i.e., $\mathcal{C} = \mathcal{C}(\mathbf{d}, \lambda, \ell)$. In view of Lemma 5.1 and the normalization property, the goal of this section is to show that if $\mathbf{d} \to \mathbf{e}_1$, then the cascade sequence does not depend on $\mathbf{d}$ anymore, i.e.

$$\mathcal{C}(\mathbf{d}, \lambda, \ell) = \mathcal{C}(\mathbf{e}_1, \lambda, \ell).$$

To analyze the power imbalance in this framework, we first introduce some notation as well as formally define the shedding rule and the redistribution of power flows.

Definition 5.1 (Uniform shedding rule)**.** *Let $\mathbf{g}^{(1)} = \mathbf{g}^*, \mathbf{d}^{(1)} = \mathbf{d}$ be the initial generation and demand vectors. Assume that the removal of lines $\ell^{(1)}, \dots, \ell^{(m)}$, $m \geq 1$, disconnects the network in components $\mathcal{G}_i^{(m)} = (\mathcal{N}_i^{(m)}, \mathcal{E}_i^{(m)})$, $i = 1, \dots, h_m$. Define the power imbalance in component $\mathcal{G}_i^{(m)}$ as*

$$Y_{\mathcal{G}_i^{(m)}} = \sum_{k \in \mathcal{N}_i^{(m)}} (g_k^{(m)} - d_k^{(m)}).$$

In order to restore power balance, generation and demand in each component are modified iteratively according to the following uniform shedding rule, for $k \in \mathcal{N}_i^{(m)}$:

$$d_k^{(m+1)} = \begin{cases} \left(1 - \frac{|Y_{\mathcal{G}_i^{(m)}}|}{\sum_{l \in \mathcal{N}_i^{(m)}} d_l^{(m)}}\right) d_k^{(m)} & \text{if } Y_{\mathcal{G}_i^{(m)}} < 0 \\ d_k^{(m)} & \text{if } Y_{\mathcal{G}_i^{(m)}} \geq 0 \end{cases},$$

$$g_k^{(m+1)} = \begin{cases} g_k^{(m)} & \text{if } Y_{\mathcal{G}_i^{(m)}} < 0 \\ \left(1 - \frac{|Y_{\mathcal{G}_i^{(m)}}|}{\sum_{l \in \mathcal{N}_i^{(m)}} g_l^{(m)}}\right) g_k^{(m)} & \text{if } Y_{\mathcal{G}_i^{(m)}} \geq 0. \end{cases}$$

Definition 5.2 (Power flow redistribution)**.** *Assume that the removal of lines $\ell^{(1)}, \dots, \ell^{(m)}, m \geq 1$, disconnects the network in components $\mathcal{G}_i^{(m)} = (\mathcal{N}_i^{(m)}, \mathcal{E}_i^{(m)})$, $i = 1, \dots, h_m$. Then, the line flows in component $\mathcal{G}_i^{(m)}$ are given by*

$$f_{\mathcal{E}_i}^{(m+1)} = \widehat{\mathbf{V}}^{(m+1, \mathcal{G}_i)} (\mathbf{g}_{\mathcal{N}_i}^{(m+1)} - \mathbf{d}_{\mathcal{N}_i}^{(m+1)}),$$

where $\widehat{\mathbf{V}}^{(m+1, G_i)}$ is the PTDF matrix for the subgraph $\mathcal{G}_i^{(m)}$, and $g_{\mathcal{N}_i}^{(m+1)}, d_{\mathcal{N}_i}^{(m+1)}$ are defined as in Definition 5.1.

In order to show the convergence of the cascade sequence, we require a second assumption.

Assumption 5.2. *For all lines j and $m \geq 2$,*

$$|f_j^{(m)}(\mathbf{e}_1)| - F_j(\mathbf{e}_1) \neq 0.$$

That is, for $\mathbf{d} = \mathbf{e}_1$ it is not possible for a line flow $|f_j^{(m)}|$ to be exactly equal to its limit. In terms of PTDF matrices and λ, this assumption reads

$$\lambda |(\widehat{\mathbf{V}}^{(m, \mathcal{G}_i)} \mathbf{e}_1)_j| \neq |(\widehat{\mathbf{V}} \mathbf{e}_1)_j|, \qquad m \geq 2.$$

Assumption 5.2 states that none of the line flows equal its emergency line limit in the cascade sequence if $d = \mathbf{e}_1$. In practice, this assumption involves excluding finitely many λ's from our analysis, which correspond to phase-transitions.

In order to prove the convergence of the cascade sequence, we also need a continuity property of the line flows at every stage with respect to the demand vector.

Lemma 5.3 (Continuity of $f_j^{(m)}$ with respect to $\mathbf{d}$)**.** *At each stage m of the cascade, the redistributed power flows $f_j^{(m)}$ are continuous in the initial demand vector $\mathbf{d}$ for all $j = 1, ..., m$.*

Proof. Assume that the removal of lines $\ell^{(1)}, \dots, \ell^{(m)}$, $m \geq 1$, disconnects the network in components $\mathcal{G}_i^{(m)} = (\mathcal{N}_i^{(m)}, \mathcal{E}_i^{(m)})$, $i = 1, \dots, h_m$. According to Definition 5.2,

$$f_{\mathcal{E}_i}^{(m+1)} = \widehat{\mathbf{V}}^{(m+1,\mathcal{G}_i)}(\mathbf{g}_{\mathcal{N}_i}^{(m+1)} - \mathbf{d}_{\mathcal{N}_i}^{(m+1)}),$$

for each connected component $\mathcal{G}_i^{(m)}$, so $\mathbf{f}^{(m+1)}$ is continuous in $\mathbf{g}^{(m+1)}, \mathbf{d}^{(m+1)}$. Moreover, according to Definition 5.1, $\mathbf{g}^{(m+1)}, \mathbf{d}^{(m+1)}$ are continuous functions of $\mathbf{g}^{(m)}, \mathbf{d}^{(m)}$. By unfolding the recursion, and using that $\mathbf{g}^*(\mathbf{d})$ is continuous in $\mathbf{d}$, we see that $\mathbf{f}^{(m+1)}$ is continuous in $\mathbf{d}$. □

Finally, we can show the main result of this section.

Proposition 5.1. *Assume that Assumptions 5.1 and 5.2 hold, and let*

$$\mathcal{C}(\mathbf{d}, \lambda, \ell) = \{\ell^{(1)}, \dots, \ell^{(T)}\}$$

be a cascade sequence initiated by $\ell = \ell^{(1)}$. Then, there exists $\varepsilon > 0$ such that

$$d_1 = 1, d_j < \varepsilon, \forall j \geq 2 \implies \mathcal{C}(\mathbf{d}, \lambda, \ell) = \mathcal{C}(\mathbf{e}_1, \lambda, \ell).$$

The proof is reported in Appendix 5.C. For an illustration on how one can easily derive the finite number of phase-transition values, we refer to [134, 170].

5.3.3 Asymptotic behavior of power imbalance

In Section 5.2.2, we showed that the only likely way to have a large blackout is when there is a single city that has a significantly larger demand than all other cities. Under certain assumptions, given the position of this city (i.e., labeling this as city 1) and the first line failure, the cascade sequence is deterministic and the same to the one as if the demand vector would have been $\mathbf{d} = \mathbf{e}_1$

(Section 5.2.3). In this section, we exploit these properties to derive the tail behavior of S, or equivalently, the amount of load that is shed/the number of affected customers.

We point out that the demands are independent and identically distributed, so the probability that a city has the largest demand equals $1/n$. To obtain the tail behavior of S, we need Assumptions 5.1 and 5.2 to hold regardless of which city has the largest power demand.

Assumption 5.3. *Assumptions 5.1 and 5.2 hold for any relabeling of the vertices.*

Note that since the number of cities n is finite, and inherently also the number of the possible lines where the first failure occurs, Assumption 5.3 excludes only a finite number of possible values of λ from our framework. The main theorem follows.

Theorem 5.1. *Suppose there is a fixed topology $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ and a fixed $\lambda \in (0, 1)$, for which Assumption 5.3 holds. Write $Z(i, \ell)$, $i = 1, ..., n$, $\ell = 1, ..., m$ as the number of cities that are not in the same component as city i after the cascade under demand vector $d = \mathbf{e}_i$ and first line failure ℓ. If $Z(i, \ell) = 0$ for all $i = 1, ..., n$ and $\ell = 1, ..., m$, then as $x \to \infty$,*

$$\mathbb{P}(S > x) = O(x^{-2\alpha}). \tag{5.24}$$

Otherwise, as $x \to \infty$, there exists a $C_1 \in (0, \infty)$ such that

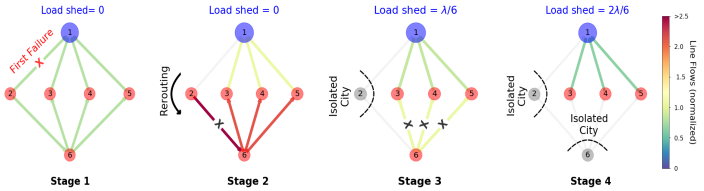
$$\mathbb{P}(S > x) \approx C_1 x^{-\alpha}. \tag{5.25}$$

The proof is reported in Appendix 5.C. See also [170].

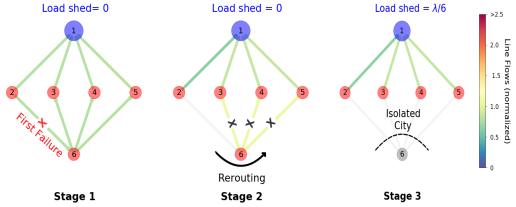
5.3.4 Blackout build-up by multiple jumps

As the cascade progresses, the size of the blackout S can gradually build up by multiple big shocks, or *jumps*, as the island containing the biggest city is shrinking in size. Large deviations theory for heavy tails can be applied to show that each of these jumps is a fixed fraction of the city size, proportional to the number of generators that are being cut off.

We illustrate this in Fig. 5.2 on a simple network of six nodes, as well as on the SciGRID simulation in Section 5.4.1, referring to [134, 170] for details. This example shows how λ acts as a tuning parameter: if $\lambda > 3/4$ the total blackout size is realized by means of either one or two big shocks (the two scenarios being equally likely) while smaller values of λ only lead to one shock. When $\lambda < 1/4$ (an overly conservative value), a single line failure does not lead to a blackout.



(a) Initial disturbance occurs at one of the upper lines.



(b) Initial disturbance occurs at one of the lower lines.

Figure 5.2: Schematic illustration of a cascade in a six-node network with $\lambda > 3/4$. The operational flows for the four lower and upper lines are $\lambda/24$ and $5\lambda/24$, respectively, with corresponding emergency line limits of $1/24$ and $5/24$. (a) The failure of one of the upper lines causes the load on the adjacent lower line to surge to $\lambda/6$, exceeding the line limit and causing this line to trip (Stage 2). This cuts off node 2 and results in a load shed of $\lambda/6$. In addition, this causes the load on the three remaining lower lines to surge to $\lambda/18$, resulting in these lines tripping as well (Stage 3). This isolates node 6 and yields a further load shed of $\lambda/6$ (Stage 4), at which point the cascade ends. Thus, the set $\mathcal{A}_1$ consists of four nodes, and a total load shed of $\lambda/3$ occurs as a result of 'two big jumps'. (b) The failure of one of the lower lines produces a load surge $\lambda/18$ on the three remaining lower lines (Stage 2), causing them to trip and cutting node 6 off. The upper lines see a reduction in load and survive, so that the set $\mathcal{A}_1$ consists of five nodes, and a total load shed of $\lambda/6$ occurs as a 'single big jump' (Stage 3).

5.4 Simulation experiments

5.4.1 SciGRID network

To support our claims further, we present experimental results using the German SciGRID network. We perform our experiments using PyPSA, a free software toolbox for power system analysis [26]. We use the dataset described in [24], which provides a model of the German electricity system based on SciGRID [118]. Data for German city sizes are obtained from [203], while the population of German districts, together with the corresponding administrative borders, are taken from [61] and [62, 53].

Since the aforementioned datasets do not include nodal demand data, we generate relative nodal demands by using population sizes and administrative borders of German NUTS3 districts, which are then rescaled with hourly nation-wide demand statistics. This procedure, based on [25], is explained in detail

below.

Nodal demand

The SciGRID model of the German power grid contains 1423 generators, 585 nodes, 489 demand nodes, and $m = 852$ transmission lines. Geographical coordinates of the demand nodes are denoted by $P_1, \dots, P_{489} \in \mathbb{R}^2$. Moreover, Germany is partitioned into 402 NUTS3 administrative districts: we denote by $\text{pop}^{\text{distr}}(j) \in \mathbb{R}$ and $\mathcal{P}_j \subseteq \mathbb{R}^2$, respectively, the population and the polygon describing the administrative borders of district j.

In order to attach the loads to the 489 demand nodes, we proceed as follows. First, we partition Germany using the Voronoi tessellation associated with the demand nodes. Since some of the nodes lie outside the border of Germany, we consider a bounding square X that contains Germany and all the P_i-s, and we define the Voronoi cells:

$$V(P_i) = \{x \in X \,:\, ||x - P_i|| \leq ||x - P_j|| \, \forall j \neq i\}.$$

Then, the population of a node P_i, denoted by $\text{pop}^{\text{node}}(i)$, is taken to be proportional to the overlapping area between $V(P_i)$ and all the NUTS3 districts that intersect $V(P_i)$. Rigorously, if we define the *transfer* matrix $\mathbf{T} \in \mathbb{R}^{489 \times 402}$ as

$$T_{i,j} = \sum_{j=1}^{402} \frac{\text{Area}\left(V(P_i) \cap \mathcal{P}_j\right)}{\text{Area}(\mathcal{P}_j)}, \tag{5.26}$$

the nodal population can be calculated as the matrix-vector product $\text{pop}^{\text{node}} = \mathbf{T}\,\text{pop}^{\text{distr}}$.

Table 5.2 and Fig. 5.4 summarizes the key statistics for the power law fits of city, district and nodal population. Fig. 5.3 shows the different partitions of Germany in NUTS3 districts and Voronoi cells associated with SciGRID demand nodes. Finally, the demand at node i at time t, denoted by $d_i(t)$, is calculated by rescaling the country-wide demand $d^{\text{germany}}(t)$ by a factor proportional to the nodal population, i.e.,

$$d_i(t) = d^{\text{germany}}(t) \cdot \frac{\text{pop}_i^{\text{node}}}{\sum_i \text{pop}_i^{\text{node}}} \tag{5.27}$$

Simulation setup

The dataset described in [24] includes hourly nodal generation time series for the entire year 2011, together with data for grid topology, lines limits, generation

Quantity	N	n_{tail}	α	$x_{\min} \cdot 10^4$	KS p-value
Cities pop.	400	271	1.29 ± 0.08	4.4 ± 1	0.35
Districts pop.	402	107	2.35 ± 0.34	22.9 ± 3.8	0.65
Nodal pop.	498	51	3.77 ± 1.07	35.7 ± 7.8	0.76

Table 5.2: PLFIT statistics for German cities, district and nodal population. The KS p-value is defined in Appendix 5.A.1.

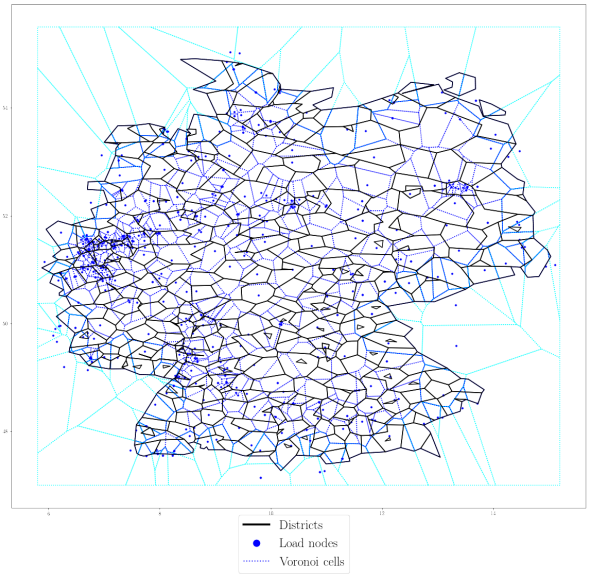


Figure 5.3: Subdivision of Germany according to NUTS3 districts and to Voronoi tessellation corresponding to demand SciGRID nodes.

capacities and marginal costs. After augmenting it with the nodal demands generated as described in Subsection 5.4.1, we can run Optimal Power Flow (OPF) instances.

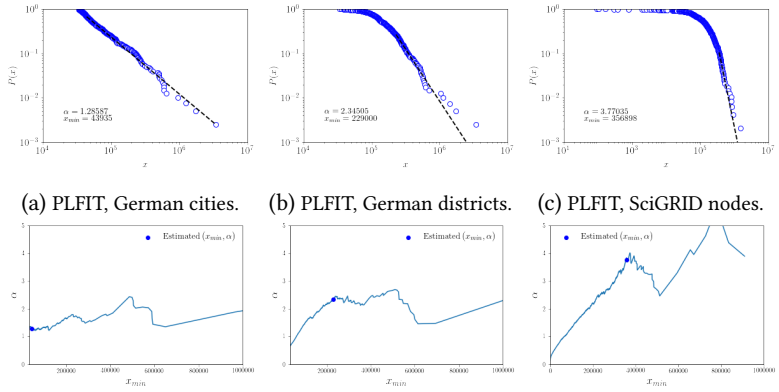
We simulate blackout data by considering one year's worth of hourly demand data. For each snapshot, we solve the operational OPF and remove one line uniformly at random, initiating a cascade. The details are reported in Algorithm 1. First, for each of the $24 \cdot 365 = 8760$ hourly snapshots of the year 2011, we solve the corresponding OPF using a safety factor $\lambda \in (0, 1)$ (line 2 in Algorithm 1). This corresponds to the *operational* phase in our mathematical model. Note that there is no *planning* phase in this simulation since we are using a model of a real-world grid.

Out of the 8760 snapshots available, only a subset results in a feasible OPF, due to the introduction on the conservative parameter λ. Such snapshots are called *feasible OPF snapshots*. Next, for each feasible snapshot, we remove one line uniformly at random (line 4), and let the cascade evolve as explained in Section 5.2 (lines 5-11). One *stage* of the cascade is comprised of lines 7-10. Note that a load shedding event (line 7) may or may not occur during a cascade stage, according to whether the previous stage line failures caused a network disconnection or not.

Finally, we store the resulting *blackout realization* (line 12) expressed in terms of the total number of customers affected, obtained from the total amount of load shed via the relationship in Eq. (5.27). In general, only a subset of the feasible snapshots resulted in *non-zero blackout realization*, i.e., a realization with a strictly positive blackout size, the others stopping without disconnecting the network, and thus without any load shedding.

Algorithm 1 Monte Carlo simulation - SciGRID German Network

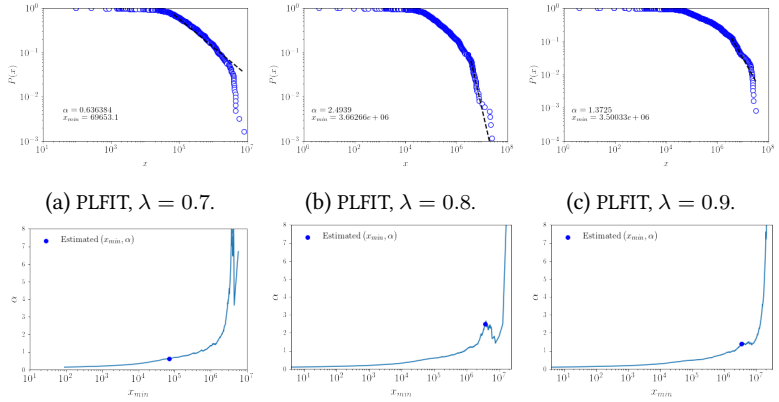
1: **Inputs:**
   $\mathcal{T} = \{$hourly snapshots for the year 2011$\}$
   $\lambda =$ line limits scaling factor
2: **Initialize:**
   Solve OPF $\forall t \in T$ with scaling factor $\lambda \in (0, 1)$
   Set $\mathcal{T}(\lambda) = \{$feasible OPF snapshots$\}$
   For all $t \in \mathcal{T}(\lambda)$, let $\mathcal{G}(t)$ be the corresponding network
3: **for** $t \in \mathcal{T}(\lambda)$ **do**
4: Remove 1 line uniformly at random from $\mathcal{G}(t)$
5: Set $\mathcal{G}(t)$.changed =True
6: **while** $\mathcal{G}(t)$.changed =True **do**
7: Shed load/generation within each component of $\mathcal{G}$ (*load shedding event*)
8: Recompute normalized power flows f_ℓ
9: Remove from $\mathcal{G}$ all lines exceeding the original line limit
10: If at least one line was removed, let $\mathcal{G}(t)$.changed =True; otherwise, set $\mathcal{G}(t)$.changed =False
11: **end while**
12: Store blackout realization
13: **end for**



(a) PLFIT, German cities. (b) PLFIT, German districts. (c) PLFIT, SciGRID nodes.

(d) Hill plot, German cities. (e) Hill plot, German districts. (f) Hill plot, SciGRID nodes.

Figure 5.4: PLFIT results for German cities, districts and nodes population.



(a) PLFIT, $\lambda = 0.7$. (b) PLFIT, $\lambda = 0.8$. (c) PLFIT, $\lambda = 0.9$.

(d) Hill plot, $\lambda = 0.7$. (e) Hill plot, $\lambda = 0.8$. (f) Hill plot, $\lambda = 0.9$.

Figure 5.5: Results for SciGRID blackout simulation for different values of λ. (a), (b), (c): PLFIT results and log-log plot of the ccdf of the number of customers affected; (d), (e), (f) Hill plots.

scaling factor	# feasible OPF snaps.	$n_{\text{non-zero}}$
0.7	3718	614
0.8	4988	858
0.9	6127	1220

Table 5.3: Key statistics for OPF feasibility and blackout realizations for the SciGRID case study. $n_{\text{non-zero}}$ denotes the number of snapshots associated to a non-zero blackout size.

5.4.2 Results and analysis

Given a cascade realization with n stages, let

$$L_i = \text{cumulative load shed up to stage } i, \quad i = 0, \dots, n,$$

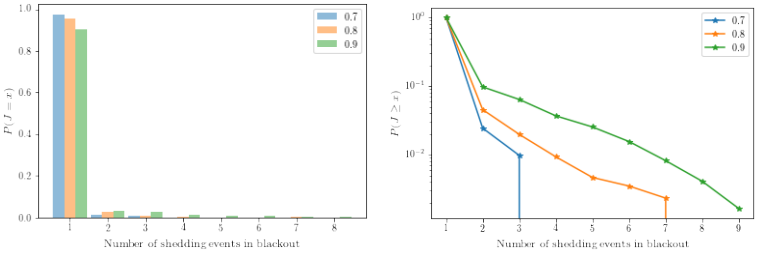
denote the cumulative amount of load shed at stage i, with the convention $L_0 = 0$, and let $L_i - L_{i-1}$ denote the amount of load shed at stage i. The *number of load shedding events*, or *jumps*, in a blackout realization with n stages is

$$J = |\{i = 1, \dots, n \,:\, L_i - L_{i-1} > 0\}|$$

Fig. 5.6 reports the histogram and the ccdf of the total number of load shedding events in the SciGRID network, for different values of λ. Uniformly across different loading factors λ, we found that the preponderance of blackouts involves just a single big load shedding event. For a moderate loading factor $\lambda = 0.7$, nearly 98% of the blackouts only involve a single jump, and even for a high loading factor $\lambda = 0.9$, 90% of the blackouts involve just a single jump, and the fraction of blackouts with four or more jumps remains below 4% in all cases, as can be appreciated from Fig. 5.6b.

Fig. 5.7 depicts the largest observed blackout across the 8760 snapshots and initial line failures, for different values of λ. Even in this massive blackout, there are only a few load shedding events, and the bulk of the load shed is the result of a single big jump. These observations sharply contrast with the branching process approximations, in which many small jumps take place.

While the single big shock jump is responsible for a large blackout, the power law behavior is not recovered (see Fig. 5.5). This is due to the small dimension of the network and the fact that German city sizes are essentially fixed, as opposed to our mathematical model where $X_1, \dots, X_n$ are random variables. For a sufficiently large network, however, a frozen version of our model still leads to the correct power law behavior, as we show in Section 5.4.3.



(a) Histogram. (b) CCDF; a log scale is used for the y-axis.

Figure 5.6: Statistics for the total number of shedding events J in the SciGRID simulation.

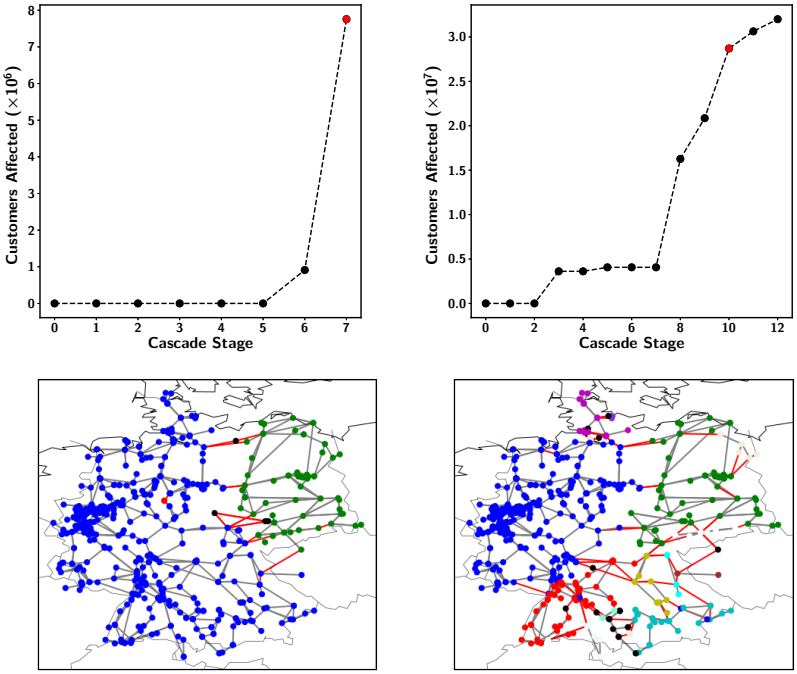


Figure 5.7: Dissection of a massive blackout in the SciGRID network for loading factors $\lambda = 0.7$ (left) and $\lambda = 0.9$ (right) in terms of the cumulative number of affected customers at each stage, as displayed in the top charts with the selected jump colored red. The islanded components corresponding to the selected jump are visualized with different colors in the bottom pictures.

5.4.3 Simulations on random graphs with frozen city sizes

Our mathematical framework described in Section 5.2 models city sizes $X_1, \dots, X_n$ as Pareto distributed random variables, while in the real world the sizes of cities served by a given power grid are essentially fixed. In this section, we show that our results still hold for a version of our model where the city sizes are kept fixed. In order to show this, we synthetically generate blackout data using a simulation based on the three-stage mathematical model described in Section 5.2. In particular, we generate a single instance of a Watts-Strogatz random graph with $n = 10000$ nodes, $m = 20000$ lines, rewiring probabilities $p \in \{0.3, 0.5, 0.7\}$ and mean degree $K = 4$, and set unitary line reactances. Each node corresponds to a city, and the city sizes are sampled from a Pareto distribution with parameters $\alpha^{(d)} = 1.37$ and $x_{\text{min}}^{(d)} = 5 \cdot 10^4$, according to the results in Table 5.1. The line limit scaling factor is set to $\lambda = 0.7$. Each iteration of the simulation stops when there are no more overloaded lines, or the graph got disconnected in two islands. We remark that the graph topology and city sizes are sampled only *once*, so that the only source of randomness during the simulation is the first outage event. The simulation setup is summarized in Algorithm 2, and the results are reported in Table 5.4 and in Fig. 5.8.

p	N_{iter}	$n_{\text{non-zero}}$	n_{tail}	α	$x_{\text{min}} \cdot 10^4$	KS p-value
0.3	2000	1046	109	1.29 ± 0.11	43.9 ± 8.0	0.5
0.5	2000	1061	277	1.45 ± 0.10	22.73 ± 5.0	0.77
0.7	2000	1111	84	1.29 ± 0.19	39.97 ± 9.0	0.57

Table 5.4: PLFIT statistics for synthetically generated blackout data, using a Watts-Strogatz random graph model for the power grid topology and keeping city sizes fixed, for different rewiring probabilities p. $n_{\text{non-zero}}$ is the number of nonzero blackout realizations. n_{tail} is the number of (nonzero) realizations x_i greater or equal than $\hat{x}_{\text{min}}$.

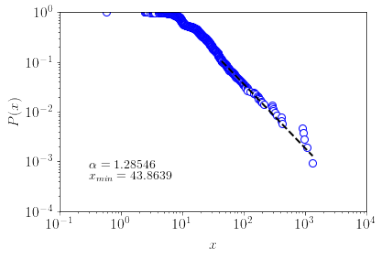
We observe that the tail index estimates α are within one standard deviation apart from the city sizes index $\alpha^{(d)} = 1.37$, with p-value ≥ 0.1, consistently across different values of the rewiring probability p. The result is corroborated by the analysis of the Hill plots in Fig. 5.8, where we observe that the flat region of the graph $x_{\text{min}} \to \alpha(x_{\text{min}})$ is close to $\alpha^{(d)} = 1.37$. We conclude that, if the network is large enough, the Pareto law of blackout sizes is inherited from that of city sizes as predicted by our model, even in the realistic case where city sizes are fixed. This is consistent with the data for the (even larger) North American grid.

Algorithm 2 Monte Carlo simulation - Synthetic datasets

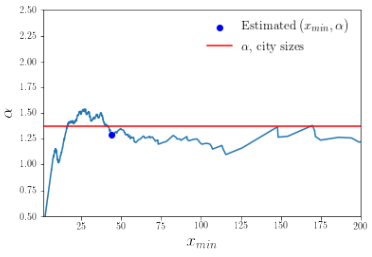
Inputs:
Parameters for sampling network and city sizes:
$n, m, p, K, \alpha^{(d)}, x_{\min}^{(d)}$
Line limit scaling parameter $\lambda \in (0, 1)$
Number of blackout realizations N_{iter}

Initialize:
Sample network topology $\mathcal{G}$ and city sizes
Solve OPF without line limits; let $\mathbf{f}^*$ be the resulting power flows
Solve OPF with line limits $\bar{\mathbf{f}} = \lambda|\mathbf{f}^*|$
Choose a random subset $\mathcal{E}'$ of lines, with cardinality $\mathcal{N}_{\text{iter}}$

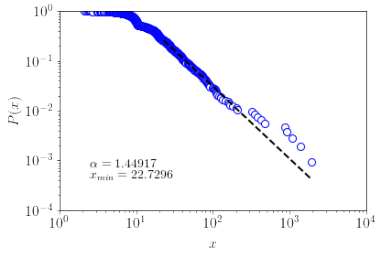
for ℓ in $\mathcal{E}'$: **do**
    Set $\mathcal{G}$.connected = True
    Remove line ℓ from $\mathcal{G}$
    Set $\mathcal{G}$.changed =True
    while $\mathcal{G}$.changed =True **and** $\mathcal{G}$.connected = True **do**
        Shed load/generation within each component of $\mathcal{G}$ to achieve balance
        Recompute power flows
        Remove from $\mathcal{G}$ the line with the largest relative overload wrt. $|\mathbf{f}^*|$
        If a line was removed, let $\mathcal{G}$.changed = True; otherwise, set $\mathcal{G}$.changed = False.
        If $\mathcal{G}$ is still connected, let $\mathcal{G}$.connected = True; otherwise, set $\mathcal{G}$.connected = False.
    end while
    Store blackout realization
end for



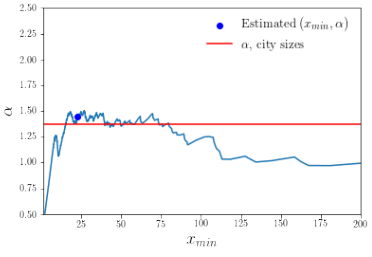
(a) $p = 0.3$.



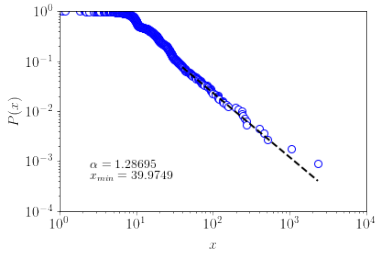
(b) $p = 0.3$.



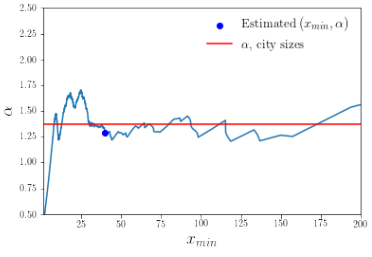
(c) $p = 0.5$.



(d) $p = 0.5$.



(e) $p = 0.7$.



(f) $p = 0.7$.

Figure 5.8: Results for synthetically generated blackout data, using a Watts-Strogatz random graph for the power grid topology and keeping city sizes fixed, for different rewiring probabilities p. a,b,c): PLFIT results and log-log plot of CCDF; d,e,f) Hill plot: red line corresponds to the city sizes tail index $\alpha^{(d)} = 1.37$.

5.5 Concluding remarks

Using data analysis, probabilistic analysis, and a simulation study, we have illustrated how extreme variations in city sizes can cause the scale-free nature of blackouts. Our explanation and refinement Eq. (5.19) of the scaling law Eq. (1.36) show that the network characteristics only appear in the pre-factor Eq. (5.19). The main parameter α, which determines how fast the probability of a big blackout vanishes as its size grows, is completely determined by the city size distribution. Decreasing the constant Eq. (5.19) by performing network upgrades (which in our framework is equivalent to decreasing λ) would only lead to a modest decrease in the likelihood of big blackouts. As a consequence, it is questionable as to whether network upgrades, as considered in [57, 206], are the most effective way to mitigate the consequences of big blackouts.

Instead, our insights suggest that it may be more effective to invest in responsive measures that enable consumers to react to big blackouts. It is shown in [86] that durations of blackouts have a tail which is decreasing much faster than (1.36). Therefore, if the goal is to minimize the negative effects of a big blackout, it may be far more effective to invest in solutions (such as local generation and storage) that aim at surviving a blackout of a specific duration. This is consistent with recent studies on the importance of resilient city design [6].

Finally, our framework and insights suggest new ways of approaching scale-free phenomena in other transportation networks: while such network topologies are not scale-free, they can still exhibit scale-free behavior, caused by the scale-free nature of nodal sizes.

Appendix

5.A Historical data analysis

5.A.1 Statistical procedure

In order to analyze the power law behavior of city and blackout sizes, we use the PLFIT method introduced in [42] to fit a Pareto distribution to a given empirical dataset $\{x_i\}_{i=1}^N$. The PLFIT method is based on a combination of the Hill estimator to find the tail exponent α, and on the Kolmogorov-Smirnov statistic to find $x_{\min}$, as outlined below. For each possible choice of $x_{\min}$, the best-fitting tail index α is found via the Hill estimator [84]

$$\hat{\alpha}(x_{\min}) = n\Big(\sum_{x_i \geq x_{\min}} \ln \frac{x_i}{x_{\min}}\Big)^{-1}.$$

Then, the KS goodness-of-fit statistic $D(x_{\min}) = \max_{x \geq x_{\min}} |S(x) - P(x)|$ is calculated, where $S(x)$ is the empirical Cumulative Distribution Function (CDF) of the data and $P(x)$ is the CDF of the Pareto distribution with parameters $x_{\min}$ and $\hat{\alpha}(x_{\min})$. Finally, the estimated $\hat{x}_{\min}$ is the one that minimizes D over all possible choices of $x_{\min}$. Uncertainty in the estimated tail exponent $\hat{\alpha}(\hat{x}_{\min})$ and lower bound $\hat{x}_{\min}$ is quantified via the nonparametric bootstrap method described in [42]. Finally, a goodness-of-fit test based on the KS statistic is used to generate a p-value that quantifies the plausibility of the power law hypothesis. The authors in [42] suggest to use the following (conservative) choice: the power law is *ruled out* if $p \leq 0.1$.

We remark that any automatic procedure for the estimation of the parameter $x_{\min}$ is imperfect and should be paired with additional, case-by-case analysis. For instance, it is not known whether the PLFIT estimator is consistent. In this chapter, we always couple the PLFIT procedure with the manual observations of the Hill plot, i.e., the graph of the mapping $x_{\min} \to \hat{\alpha}(x_{\min})$, and report whether the PLFIT results are consistent with the visual analysis of this plot, i.e., whether $\hat{x}_{\min}$ lies within a region where the values of α are relatively stable.

5.A.2 Data pre-processing and results

In this section, we analyze the scale-free behavior of US city and blackout sizes. The data for US city sizes, as per the 2000 US census, are available in [42]. The data for US blackouts are extracted from the Electric Disturbance Events Annual Summaries, Form OE-417 [184] of the US Department of Energy, which covers the period 2002-2018. Each record of the OE-417 dataset contains information on the date, area of interest and number of customers affected in a single outage event. Here, the size of a blackout is defined as the number of customers affected by it.

The presence of missing or noisy records in the dataset requires the following pre-processing actions: i) records for which the "Number of customers affected" entry is unknown are removed; ii) records for which the "Number of customers affected" consists of two or more values, corresponding to different US states, are modified by replacing the multiple values with their sum; iii) records for which the "Number of customers affected" entry is not purely numeric are removed. The only two exceptions to iii) are when *both* the "cumulative" and "peak" number of customers affected are reported (in which case only the "cumulative" values is retained), and when the number of customers affected is described by a range of values (in which case the midpoint value is retained).

Fig. 5.1 reports the PLFIT results and the corresponding Hill plots. We observe that the estimated parameters lie in the flat portion of the Hill plots.

5.B Properties of $\widehat{\mathbf{V}}$ and power flow redistribution

The following lemma is based on a well-known result in graph theory (see, for example, [185]).

Lemma 5.4. *If $\mathcal{G}$ is a connected graph, rk $(\widehat{\mathbf{V}}) =$ rk $(\mathbf{A}) =$ rk $(\mathbf{L}) =$ rk $(\mathbf{L}^+) = n-1$, and the null space of $\widehat{\mathbf{V}}$ is the one-dimensional subspace generated by $\mathbf{e} = (1, \ldots, 1) \in \mathbb{R}^n$, i.e.*

$$Ker\,(\widehat{\mathbf{V}}) = Ker\,(\mathbf{A}) = Ker\,(\mathbf{L}) = Ker\,(\mathbf{L}^+) =< \mathbf{e} > .$$

Recall that the matrix $\widehat{\mathbf{V}}$ depends on the specific orientation that has been chosen for the edges $\mathcal{E}$ (Section 1.2.2). The following lemmas describe the dependency of $\widehat{\mathbf{V}}$ on the chosen orientation in more detail.

Lemma 5.5. *Changing the orientation of a subset of lines $\mathcal{E}' \subset \mathcal{E}$ has the effect of swapping the sign of the corresponding rows of the PTDF matrix $\widehat{\mathbf{V}}$. In particular, it is always possible to choose the orientation such that $\widehat{\mathbf{V}}\mathbf{e}_1 \geq 0$.*

Proof. Changing the orientation of a line from $l_k = (i, j)$ to $\tilde{l}_k = (j, i)$, by definition, amounts to swapping the sign of the k-th row of matrix $\mathbf{A}$, yielding a modified matrix $\overline{\mathbf{A}} = \mathbf{I}^{(k)}\mathbf{A}$, where $\mathbf{I}^{(k)}$ is a diagonal matrix with $I^{(k)}_{ii} = 1$ if $i \neq k$ and $I^{(k)}_{kk} = -1$. Since $\tilde{\mathbf{L}} = \overline{\mathbf{A}}^\top \overline{\mathbf{A}} = \mathbf{A}^\top \mathbf{I}^{(k)} \mathbf{I}^{(k)} \mathbf{A} = \mathbf{A}^\top \mathbf{A} = \mathbf{L}$, the matrices $\mathbf{L}$ and $\mathbf{L}^+$ are not affected by the change. As a consequence, the modified PTDF matrix $\overline{\widehat{\mathbf{V}}} = \mathbf{D}\overline{\mathbf{A}}\mathbf{L}^+ = \mathbf{I}^{(k)}\widehat{\mathbf{V}}$ differs from $\widehat{\mathbf{V}}$ only by the swapped signs on the k-th row. □

Lemma 5.6. *Let $\mathcal{G}$ be assigned the orientation such that the set of edges incident to node 1 is $\mathcal{E}_1 = \{(1, j) \,|\, j \text{ is adjacent to } 1\}$, i.e. $A_{\ell,1} = 1 = -A_{\ell,j}$ for all $\ell = (1, j) \in \mathcal{E}_1$. Then, $V_{\ell,1} \geq 0$ for every $\ell \in \mathcal{E}_1$. The converse is also true.*

Proof. First, note that largest element in each row of L^+ is its diagonal entry (Corollary 1 in [186]), i.e. $L^+_{1,1} - L^+_{1,j} \geq 0$ for every $\ell = (1, j) \in \mathcal{E}_1$. For any line $\ell = (1, j) \in \mathcal{E}_1$, we have $V_{\ell,1} = (\mathbf{A}\mathbf{L}^+)_{\ell,1} = A_{\ell,1}L^+_{1,1} + A_{\ell,j}L^+_{1,j}$, where $A_{\ell,1} = -A_{\ell,j} = \pm 1$ depending on the orientation of line ℓ. Thus, $V_{\ell,1} \geq 0$ if and only if $A_{\ell,1} = 1 = -A_{\ell,j}$. □

In the event of the failure of a subset of transmission lines $\mathcal{E}' \subset \mathcal{E}$, and provided that the power injections remain unchanged, the power flows will redistribute among the remaining lines according to power flow physics, provided that the altered graph $\widetilde{G} = (\mathcal{N}, \mathcal{E} \setminus \mathcal{E}')$ remained connected. The way the

power flows redistribute is governed by the new PTDF matrix $\widetilde{\widehat{\mathbf{V}}}$, which can be constructed analogously to $\widehat{\mathbf{V}}$, mapping the (unchanged) power injections to the new power flows. We assume that the redistribution occurs instantaneously, without any transient effects.

As an illustration, we show how the redistributed power flows can be calculated in the special case of an isolated failure $\mathcal{E}' = \{\ell\}$. In this case, it is enough to calculate the vector $\boldsymbol{\phi}^{(\ell)} \in \mathbb{R}^{m-1}$ of redistribution coefficients, known as *line outage distribution factors*. The quantity $\phi_j^{(\ell)}$ takes values in $[-1, 1]$, and $|\phi_j^{(\ell)}|$ represents the percentage of power flowing in line ℓ that is redirected to line j after the failure of the former. In particular, the new power flow configuration after the failure of line $\ell = (i, j)$, denoted by $\mathbf{f}^{(\ell)} \in \mathbb{R}^{m-1}$, is given by

$$f_k^{(\ell)} = f_k + f_\ell^{(\ell)} \phi_k^{(\ell)}, \; \forall \ell \neq k, \tag{5.28}$$

where, for $k = (a, b)$ and $\ell = (i, j)$, the coefficient $\phi_{k,\ell} \in \mathbb{R}$ can be computed as

$$\phi_{k,\ell} = \phi_{(a,b),(i,j)} = x_{a,b}^{-1} \cdot \frac{R_{a,j} - R_{a,i} + R_{b,i} - R_{b,j}}{2(1 - x_{i,j}^{-1} R_{i,j})}, \tag{5.29}$$

where $R_{i,j}$ is the *effective resistance* between nodes i and j, given by

$$R_{i,j} = (\mathbf{e}_i - \mathbf{e}_j)^T L^+ (\mathbf{e}_i - \mathbf{e}_j) = (L^+)_{i,i} + (L^+)_{j,j} - 2(L^+)_{i,j}.$$

.

5.C Extended proofs

Proof of Lemma 5.1. We observe that the total mismatch can never exceed the sum of all demands, and hence

$$S \leq \sum_{i=1}^{n} d_i \leq n d_1.$$

Therefore,

$$\begin{aligned}
&\mathbb{P}\left(S > x; d_i > \sum_{i=2}^{n} d_i \geq \epsilon d_1 \right) \\
\leq &\mathbb{P}\left(d_1 > \frac{x}{n}; d_i > \epsilon \frac{d_1}{n} \text{ for some } i = 2, ..., n \right) \\
\leq &\mathbb{P}\left(d_i > \epsilon \frac{x}{n^2} \text{ for some } i = 2, ..., n \right).
\end{aligned}$$

Write $I(y) = |\{i : d_i > y\}|$. Since for every $\eta > 0$,

$$\mathbb{P}\left(I(\eta x) \geq 2\right) = O\left(x^{-2\alpha}\right)$$

as $x \to \infty$, the result follows. □

Proof of Lemma 5.2. First, note that without loss of generality, we can always normalize our framework by dividing all parameters (e.g. generation, line limits, etc.) by the sum of all power demands. This yields an equivalent setting where the total power demand equals one, i.e., $\mathbf{d} = \mathbf{e}_1$. Therefore, we will normalize $y = 1$ in the derivation below.

First, we note that the selected orientation on $\mathcal{G}$ implies that the set of edges incident to node 1 is $\mathcal{E}_1 = \{(1, j) \,|\, j \text{ is adjacent to } 1\}$ (i.e., the edges in $\mathcal{E}_1$ *exit* node 1), or, in terms of the edge-node incidence matrix $\mathbf{A}$, that $A_{\ell,1} = 1 = -A_{\ell,j}$ for all $\ell = (1, j) \in \mathcal{E}_1$. This is proved in Lemma 5.6 in Appendix 5.B. Due to the chosen orientation, $\overline{\mathbf{f}}(\lambda) = \lambda\,|\widehat{\mathbf{V}}\mathbf{e}_1| = \lambda\widehat{\mathbf{V}}\mathbf{e}_1$ and the line limit constraints in $P(\lambda)$ can be rewritten as

$$(1-\lambda)\widehat{\mathbf{V}}\mathbf{e}_1 \leq \widehat{\mathbf{V}}\mathbf{g} \leq (1+\lambda)\widehat{\mathbf{V}}\mathbf{e}_1.$$

The problem $P(\lambda)$ is a strictly convex optimization problem with linear equality and inequality constraints. Therefore, in order to show that $\mathbf{g}(\lambda)$ is the unique optimal solution, it is sufficient to show that it satisfies the KKT conditions for $P(\lambda)$, which read

$$\mathbf{g} + \widehat{\mathbf{V}}^\top(\boldsymbol{\mu}^+ - \boldsymbol{\mu}^-) + \gamma\mathbf{e} = 0, \tag{5.30}$$

$$\boldsymbol{\mu}^+ \geq \mathbf{0}, \boldsymbol{\mu}^- \geq \mathbf{0}, , \tag{5.31}$$

$$\mu_l^+(\widehat{\mathbf{V}}\mathbf{g} - (1+\lambda)\widehat{\mathbf{V}}\mathbf{e}_1)_\ell = 0\,\forall \ell \in \mathcal{E}, \tag{5.32}$$

$$\mu_l^-(-\widehat{\mathbf{V}}\mathbf{g} + (1-\lambda)\widehat{\mathbf{V}}\mathbf{e}_1)_\ell = 0\,\forall \ell \in \mathcal{E}, \tag{5.33}$$

$$\mathbf{e}^\top\mathbf{g} = 1, \tag{5.34}$$

$$(1-\lambda)\widehat{\mathbf{V}}\mathbf{e}_1 \leq \widehat{\mathbf{V}}\mathbf{g} \leq (1+\lambda)\widehat{\mathbf{V}}\mathbf{e}_1, \tag{5.35}$$

where γ is the Lagrange multipliers for the equality constraint and $\boldsymbol{\mu}^+, \boldsymbol{\mu}^- \in \mathbb{R}^m$ are the Lagrange multipliers for the inequality constraints.

Since $\widehat{\mathbf{V}}\mathbf{g}(\lambda) = (1-\lambda)\mathbf{V}\mathbf{e}_1$ and $\mathbf{e}^\top\mathbf{g}(\lambda) = \mathbf{e}^\top\mathbf{e}_1 = 1$, the candidate solution $\mathbf{g}(\lambda)$ clearly satisfies the feasibility conditions (5.34) - (5.35) and the complementary slackness condition (5.33). Moreover, condition (5.32) is satisfied if we choose $\boldsymbol{\mu}^+ = \mathbf{0}$.

Using the facts that $\widehat{\mathbf{V}}\mathbf{g}(\lambda) = (1-\lambda)\widehat{\mathbf{V}}\mathbf{e}_1$ and $\text{Ker}(\widehat{\mathbf{V}}) =< \mathbf{e} >$, pre-multiplying Eq. (5.30) by $\widehat{\mathbf{V}}$ yields $(1-\lambda)\mathbf{e}_1 + \widehat{\mathbf{V}}^\top\boldsymbol{\mu} \in \text{Ker}(\widehat{\mathbf{V}})$. This is equivalent

to

$$(1-\lambda)\mathbf{e}_1 + \widehat{\mathbf{V}}^\top \boldsymbol{\mu} = \frac{(1-\lambda)}{n}\mathbf{e},$$

where in the last equality we used again the property that $\widehat{\mathbf{V}}\mathbf{e} = \mathbf{0}$. To conclude the proof, it remains to be shown that that there exist a nonnegative solution $\boldsymbol{\mu}^- \geq \mathbf{0}$ of the matrix equation

$$\widehat{\mathbf{V}}^\top(-\boldsymbol{\mu}^-) = (1-\lambda)(\mathbf{e}/n - \mathbf{e}_1). \tag{5.36}$$

We construct a non-negative solution $\boldsymbol{\mu}^-$ as follows:

$$\mu_\ell^- := (1-\lambda)\mathbf{e}_{\mathcal{E}_1} = \begin{cases} (1-\lambda) & l \in \mathcal{E}_1 \\ 0 & l \notin \mathcal{E}_1, \end{cases}$$

where $\mathbf{e}_{\mathcal{E}_1}$ is a m- dimensional vector containing ones in positions given by $\mathcal{E}_1$, and 0 elsewhere. Invoking Lemma 5.6 we see that $\mathbf{A}\mathbf{e}_1 = \mathbf{e}_{\mathcal{E}_1}$, yielding $\boldsymbol{\mu}^- = (1-\lambda)\mathbf{A}\mathbf{e}_1$. Using the definition of $\widehat{\mathbf{V}} = \mathbf{A}\mathbf{L}^+, \mathbf{L} = \mathbf{A}^\top\mathbf{A}$, and the property $\mathbf{L}^+\mathbf{L} = (\mathbf{I} - \mathbf{J}/n)$ (see [186]), we observe that Eq. (5.36) is indeed satisfied:

$$\begin{aligned}\widehat{\mathbf{V}}^\top\boldsymbol{\mu}^- &= -(1-\lambda)\widehat{\mathbf{V}}^\top\boldsymbol{\mu}^- = -(1-\lambda)\widehat{\mathbf{V}}^\top\mathbf{A}\mathbf{e}_1 \\ &= -(1-\lambda)(\mathbf{L}^+\mathbf{L})\mathbf{e}_1 = (1-\lambda)(\mathbf{e}/n - \mathbf{e}_1).\end{aligned}$$

Setting $\gamma = -1/n$ completes the proof. □

Proof of Proposition 5.1. Let $\ell^{(1)}$ be the first failure, and consider

$$\ell^{(2)} = \underset{j\in\mathcal{A}^{(2)}}{\arg\max}\Big\{\frac{|f_j^{(2)}|}{F_j}\Big\},$$

where $\mathcal{A}^{(2)} = \mathcal{A}^{(2)}(\mathbf{d},\lambda) = \{j : |f_j^{(2)}| \geq F_j\}$. Lemma 5.3 shows that $f_j^{(2)}(\mathbf{d}) \to f_j^{(2)}(\mathbf{e}_1)$ as $\mathbf{d} \to \mathbf{e}_1$, so by continuity and Assumption 5.2

$$|f_j^{(2)}(\mathbf{d})| - F_j(\mathbf{d}) \to |f_j^{(2)}(\mathbf{e}_1)| - F_j(\mathbf{e}_1) \neq 0.$$

Consequently, there exist $\varepsilon > 0$ such that, if $d_j < \varepsilon$ for all $j \geq 2$, then

$$|f_j^{(2)}(\mathbf{d})| > F_j(\mathbf{d}) \iff |f_j^{(2)}(\mathbf{e}_1)| > F_j(\mathbf{e}_1).$$

In other words, a line limits is exceeded for $\mathbf{d} = \mathbf{e}_1$ (which, due to our assumption, implies that it is strictly exceeded) if and only if it is also (strictly) exceeded when $\mathbf{d}$ is close enough to $\mathbf{e}_1$, implying that $\mathcal{A}^{(2)}(\mathbf{d}) = \mathcal{A}^{(2)}(\mathbf{e}_1)$.

Moreover, there exists $\varepsilon^{(1)} \leq \varepsilon$ such that, if $d_k < \varepsilon$ for $k \geq 2$, then

$$\max_{j \in \mathcal{A}^{(2)}(\mathbf{d}_1, \lambda)} \frac{|f_j^{(2)}(\mathbf{d}, \lambda)|}{F_j(\mathbf{d})} = \max_{j \in \mathcal{A}^{(2)}(\mathbf{e}_1, \lambda)} \frac{|f_j^{(2)}(\mathbf{d}, \lambda)|}{F_j(\mathbf{d})} = \max_{j \in \mathcal{A}^{(2)}(\mathbf{e}_1, \lambda)} \frac{|f_j^{(2)}(\mathbf{e}_1, \lambda)|}{F_j(\mathbf{e}_1)},$$

where in the first equality we used that $\mathcal{A}^{(2)}(\mathbf{d}) = \mathcal{A}^{(2)}(\mathbf{e}_1)$, and in the second the continuity property. Finally, Assumption 5.1 allows us to conclude that the max is unique and that the (unique) second failure $\ell^{(2)}(\mathbf{d}, \lambda) = \ell^{(2)}(\mathbf{e}_1, \lambda)$ does not depend on $\mathbf{d}$ if $d_k < \varepsilon^{(1)}$, $k \geq 2$.

As Lemma 5.3 holds for every stage of the cascade, we can repeat the steps above to construct a sequence $\varepsilon^{(T)} \leq \ldots, \varepsilon^{(2)} \leq \varepsilon^{(1)}$ such that the cascade sequence $\mathcal{C}$ is well defined and does not depend on $\mathbf{d}$ if $d_j < \varepsilon^{(T)}$ for all $j \geq 2$. □

Proof of Theorem 5.1. First, since the demands are independent and identically distributed, we observe that each city has an equal probability of being the city with the largest demand. That is, if B denotes the city that has the largest demand, then

$$\mathbb{P}(B = i) = 1/n, \qquad i = 1, ..., n.$$

By the law of total probability,

$$\mathbb{P}\left(S > x\right) = \sum_{i=1}^{n} \frac{1}{n} \mathbb{P}\left(S > x \,\middle|\, B = i\right).$$

Fix some $\epsilon > 0$ (sufficiently small), and note that for all $i = 1, ...n$,

$$\begin{aligned}\mathbb{P}\left(S > x \,\middle|\, B = i\right) =& \mathbb{P}\left(S > x; \sum_{j \neq i}^{n} d_j < \epsilon d_i \,\middle|\, B = i\right) \\ &+ \mathbb{P}\left(S > x; \sum_{j \neq i}^{n} d_j \geq \epsilon d_i \,\middle|\, B = i\right).\end{aligned}$$

Due to Lemma 5.1, we observe that the second term is of order $O(x^{-2\alpha})$ for all $i = 1, ..., n$, and, hence,

$$\sum_{i=1}^{n} \frac{1}{n} \mathbb{P}\left(S > x; \sum_{j \neq i}^{n} d_j \geq \epsilon d_i \,\middle|\, B = i\right) = O(x^{-2\alpha}).$$

For the first term, note that Assumption 5.3 ensures that $Z(i,\ell)$ is well-defined for all $i = 1, ..., n$ and $\ell = 1, ..., m$. Since we choose our first failure uniformly at random among all lines, we observe that by law of total probability, for all $i = 1, ..., n$,

$$\begin{aligned}&\mathbb{P}\left(S > x; \sum_{j\neq i}^{n} d_j < \epsilon d_i \,\middle|\, B = i\right)\\ &= \sum_{l=1}^{m} \frac{1}{m}\mathbb{P}\left(S > x; \sum_{j\neq i}^{n} d_j < \epsilon d_i \,\middle|\, \ell^{(1)} = l, B = i\right).\end{aligned}$$

In case that $Z(i,\ell) = 0$ for all $i = 1, ..., n$ and $\ell = 1, ..., m$, it follows from Proposition 5.1 that for all $\epsilon > 0$ sufficiently small, the cascade sequence causes no disconnections for every city with largest demand and first line failure $\ell^{(1)}$. That is, for all $i = 1, ..., n$, $l = 1, ..., m$, $x > 0$ and $\epsilon > 0$ sufficiently small,

$$\mathbb{P}\left(S > x; \sum_{j\neq i}^{n} d_j < \epsilon d_i \,\middle|\, \ell^{(1)} = l, B = i\right) = 0.$$

Therefore, if $Z(i,\ell) = 0$ for all $i = 1, ..., n$ and $\ell = 1, ..., m$, then for all $\epsilon > 0$ sufficiently small,

$$\sum_{i=1}^{n} \frac{1}{n}\mathbb{P}\left(S > x; \sum_{j\neq i}^{n} d_j < \epsilon d_i \,\middle|\, B = i\right) = 0,$$

and we conclude that (5.24) holds.

Next, suppose that $Z(i,\ell) \neq 0$ for at least some $i \in \{1, ..., n\}$ and $\ell \in \{1, ..., m\}$. It follows from Proposition 5.1 that for all $i \in \{1, ..., n\}$ and $\ell \in \{1, ..., m\}$ for which $Z(i,\ell) \neq 0$, it holds for all $\epsilon > 0$ sufficiently small that the cascade sequence is the same as the one when the demand vector would have been $\mathbf{d} = \mathbf{e}_i$. In particular, whenever $\sum_{j\neq i}^{n} d_j < \epsilon d_i$, it holds for all $\epsilon > 0$ sufficiently small that the set A_1 is deterministic and is the same set of nodes as if demand would have been $\mathbf{d} = \mathbf{e}_i$, and the number of cities disconnected from city i equals $Z(i,\ell)$. Recall Lemma 5.2 and the property that the generator vector $\mathbf{g}$ is a continuous function of $\mathbf{d}$. Consequently, for all $i \in \{1, ..., n\}$ and $\ell \in \{1, ..., m\}$ for which it holds that $Z(i,\ell) \neq 0$, and for all $\epsilon > 0$ sufficiently

small,

$$\begin{aligned}
&\mathbb{P}\left(S > x, \sum_{j\neq i}^{n} d_j < \epsilon d_i \,\middle|\, \ell^{(1)} = \ell, B = i\right) \\
=&\mathbb{P}\left(\sum_{i\notin A_1} (g_i - d_i) > x, \sum_{j\neq i}^{n} d_j < \epsilon d_i \,\middle|\, \ell^{(1)} = \ell, B = i\right) \\
\leq&\mathbb{P}\left(Z(i,\ell)\left(\frac{\lambda}{n} + c_1(\epsilon)\right) d_i > x, \sum_{j\neq i}^{n} d_j < \epsilon d_i \,\middle|\, \ell^{(1)} = \ell, B = i\right),
\end{aligned}$$

where $c_1(\epsilon)$ is a strictly positive function with $c_1(\epsilon) \to 0$ as $\epsilon \downarrow 0$. For independent identically Pareto-distributed random variables $X_1, ..., X_n$, it holds that as $x \to \infty$,

$$\mathbb{P}\left(\max\{X_1, ..., X_n\} \geq x\right) \approx n\mathbb{P}(X_i > x) = nKx^{-\alpha}.$$

Therefore, for all $i \in \{1, ..., n\}$ and $\ell \in \{1, ..., m\}$ for which $Z(i,\ell) \neq 0$,

$$\begin{aligned}
&\lim_{\epsilon\downarrow 0}\lim_{x\to\infty} x^{\alpha}\mathbb{P}\left(S > x, \sum_{j\neq i}^{n} d_j < \epsilon d_i \,\middle|\, \ell^{(1)} = \ell, B = i\right) \\
\leq&\lim_{\epsilon\downarrow 0} n\left(Z(i,\ell)\left(\frac{\lambda}{n} + c_1(\epsilon)\right)\right)^{\alpha} \\
=&nK\left(Z(i,\ell)\frac{\lambda}{n}\right)^{\alpha}.
\end{aligned}$$

Similarly, we can obtain the same lower bound, i.e.

$$\begin{aligned}
&\lim_{\epsilon\downarrow 0}\lim_{x\to\infty} x^{\alpha}\mathbb{P}\left(S > x, \sum_{j\neq i}^{n} d_j < \epsilon d_i \,\middle|\, \ell^{(1)} = \ell, B = i\right) \\
\geq&nK\left(Z(i,\ell)\frac{\lambda}{n}\right)^{\alpha}.
\end{aligned}$$

We conclude that, as $x \to \infty$,

$$\mathbb{P}\left(S > x\right) = \sum_{i=1}^{n}\sum_{l=1}^{m} \frac{K}{m}\left(Z(i,\ell)\frac{\lambda}{n}\right)^{\alpha} x^{-\alpha}.$$

Note that term in front of $x^{-\alpha}$ is a double sum of finitely many terms, and hence we can also conclude that (5.25) holds. □

Chapter 6

A Structured Learning Approach to Predicting Locational Marginal Prices

Contents

As discussed in Section 1.3.2, fluctuations in renewable energy production and demand introduce variability in electricity prices. In this chapter, we develop a machine learning methodology to predict electricity prices in wholesale energy markets adopting the Locational Marginal Pricing (LMP) architecture described in Section 1.3.2. Our approach takes a decentralized perspective and uses only publicly available data, which are limited to historical aggregated grid-wide demand and supply, and historical nodal prices. Conversely, they do not include information such as grid topology, line limits, and nodal information about demand and supply.

The decentralized perspective makes this problem particularly challenging, since the lack of information on power grids parameters precludes the possibility of directly solving OPF problems, which are used to calculate LMPs as functions of demand and available supply. Our methodology overcomes this challenge by *learning* such functions using techniques from machine learning, convex optimization, and multiparametric programming theory. In particular, we exploit structural characteristics of the OPF mechanism to characterize LMPs as piecewise affine functions of nodal demand and renewable supply, which are learned using convex optimization and machine learning techniques that leverage sparsity properties of power grids. Finally, LMPs are predicted based on forecasted grid-wide demand and renewable supply.

In spite of the limitations inherent to the decentralized perspective, our methodology performs remarkably well in forecasting price fluctuations, and is validated using the IEEE 30-bus test network. This chapter is based on the paper [154], which includes an extension of this work that includes a validation on real market data from the Southwest Power Pool market.

Chapter outline: The chapter is organized as follows. In Section 6.1 we describe the system model and present our main contributions, while Section 6.2 contains the necessary background on multiparametric programming theory. Next, Section 6.3 provides an overview of the forecasting methodology and describe its main components, which are then illustrated in more detail in Sections 6.4 and 6.5. Numerical results on the IEEE 30-bus test case are presented in Section 6.6, while Section 6.7 provides some concluding remarks.

6.1 Model and problem formulation

6.1.1 Power system model

The power grid model is based on the description in Section 1.2.2, which we briefly recall below. The transmission network is modeled as a connected graph $\mathcal{G} = \mathcal{G}(\mathcal{N}, \mathcal{E})$, where the set of nodes $\mathcal{N} = \{1, \dots, n\}$ represents the n buses in the system, and the set of edges $\mathcal{E}$ model the m transmission lines. We assume that $\mathcal{N} = \mathcal{N}_g \sqcup \mathcal{N}_w$, with $|\mathcal{N}_g| = n_g, |\mathcal{N}_w| = n_w$, $n_g + n_w = n$, and where $\sqcup$ denotes a disjoint union. Each node in $\mathcal{N}_g$ houses a traditional controllable generator, while each node in $\mathcal{N}_w$ houses a renewable generator, like a wind or a solar farm. Without loss of generality, we assume that $\mathcal{N}_g = \{1, \dots, n_g\}$ and $\mathcal{N}_w = \{n_g + 1, \dots, n\}$. Furthermore, we let a subset of nodes $\mathcal{N}_d \subseteq \mathcal{N}$ house loads, with $|\mathcal{N}_d| = n_d$.

We denote the vectors of generations and demand, respectively, by $\mathbf{g} \in \mathbb{R}_+^{\mathcal{N}_g}$

and $\mathbf{d} \in \mathbb{R}_+^{\mathcal{N}_d}$, where the notation $x \in \mathbb{R}^A$ indicates that the entries in the $|A|$-dimensional vector x are indexed by the set A. To simplify notation, we extend the vector $\mathbf{d}$ to a n-dimensional vector $\tilde{\mathbf{d}} \in \mathbb{R}^n$ by setting $\tilde{d}_i = 0$ whenever $i \notin \mathcal{N}_d$, and $\tilde{d}_i = d_i$ otherwise.

The starting point of our methodology is the Optimal Power Flow (OPF) problem, which has been introduced in Section 1.3.1. To optimally match power demand and supply, while satisfying the power grid operating constraints, the Independent System Operator (ISO) solves an OPF problem and calculates the optimal energy dispatch vector $\mathbf{g}^* \in \mathbb{R}^n$, as well as the vector of nodal prices $\mathbf{LMP} \in \mathbb{R}^n$ (see Section 6.1.2).

In its full generality, the OPF is a nonlinear and nonconvex optimization problem, which is difficult to solve [126, 17]. For the purpose of this chapter, we focus on the widely used approximation of the OPF problem known as DC-OPF (see Section 1.3.1). Following standard practice[72, 177], for every $i \in \mathcal{N}$ the generation cost function $J_i(\cdot)$ at node i is modeled as an increasing quadratic function, resulting in the following quadratic optimization problem:

$$\min_{\mathbf{g}\in\mathbb{R}^n} \sum_{i=1}^{n} J_i(g_i) = \frac{1}{2}\mathbf{g}^\top \mathbf{H}\mathbf{g} + \mathbf{h}^\top \mathbf{g} \tag{6.1}$$

$$\text{s.t.} \quad \mathbf{1}^\top(\mathbf{g} - \tilde{\mathbf{d}}) = 0 \qquad : \lambda_{\text{en}} \tag{6.2}$$

$$\underline{\mathbf{f}} \le \widehat{\mathbf{V}}(\mathbf{g} - \tilde{\mathbf{d}}) \le \bar{\mathbf{f}} \qquad : \boldsymbol{\mu}^-, \boldsymbol{\mu}^+ \tag{6.3}$$

$$\underline{\mathbf{g}} \le \mathbf{g} \le \bar{\mathbf{g}} \qquad : \boldsymbol{\tau}^-, \boldsymbol{\tau}^+, \tag{6.4}$$

where the variables are defined as follows:

$\mathbf{H} \in \mathbb{R}^{n\times n}$	diagonal positive definite matrix defining the quadratic part of the objective function;
$\mathbf{h} \in \mathbb{R}^n$	n-dimensional vector defining the linear part of the objective function;
$\widehat{\mathbf{V}} \in \mathbb{R}^{m\times n}$	PTDF matrix (see Section 1.2.2, Eq. (1.20));
$\underline{\mathbf{f}}, \bar{\mathbf{f}} \in \mathbb{R}^m$	vector of lower/upper transmission line limits;
$\underline{\mathbf{g}}, \bar{\mathbf{g}} \in \mathbb{R}^n$	vector of lower/upper generation constraints;
$\lambda_{\text{en}} \in \mathbb{R}$	dual variable of the energy balance constraint;
$\boldsymbol{\mu}^-, \boldsymbol{\mu}^+ \in \mathbb{R}_+^m$	dual variables of the transmission line constraints (6.3),
$\boldsymbol{\tau}^-, \boldsymbol{\tau}^+ \in \mathbb{R}_+^n$	dual variables of the generation constraints (6.4);
$\mathbf{1} \in \mathbb{R}^n$	is a n-dimensional vector of ones.

In what follows, we denote by $\mathbf{J}(\mathbf{g}) := \sum_{i=1}^n J_i(g_i)$ the objective function, and we assume that $\underline{g}_i = 0$ for all renewable generators $i \in \mathcal{N}_w$. The matrix $\widehat{\mathbf{V}}$ describes the linear mapping from nodal power injections to active power

flows over transmission lines under the assumption of the DC-approximation, and reads

$$\widehat{\mathbf{V}} = [\mathbf{0}\ \mathbf{D}\widetilde{\mathbf{A}}\widetilde{\mathbf{L}}^{-1}], \tag{6.5}$$

where the matrices $\mathbf{D}, \widetilde{\mathbf{A}}, \widetilde{\mathbf{L}}$ describe topological and physical properties of the grid, and have been defined in Section 1.2.2. For this chapter, we have chosen the formulation in Eq. (1.20) in order to be consistent with the framework of [101], as some of the results obtained therein will be used in the remainder of this chapter.

6.1.2 Locational marginal prices

In this chapter and in the next Chapter 7, we focus on wholesale energy markets adopting the concept of Locational Marginal Prices (LMPs) as electricity prices at the grid nodes. Under this market architecture, which is widely used in the U.S. (see the relevant discussion in Section 1.4.3), the energy price at a specific node is defined as the marginal cost of supplying the next increment of load at that node, consistent with all power grid operating constraints. LMP-markets are usually divided into day-ahead (DA) and real-time (RT) markets. In the DA market, participants submit bids/offers to buy/sell energy. The ISO then runs the OPF to derive day-ahead LMPs for each grid's node, together with the optimal scheduled generation dispatch $\mathbf{g}^*$. Since day-ahead scheduled supply may not meet real-time demand, ISOs also calculate real-time LMPs as often as every five minutes [33].

Recall that the LMP at a specific bus is defined as the least cost to service the next increment of demand at that location consistent with all power system operating constraints, and in the case of the DC-OPF in Eqs. (6.1) - (6.4) the LMP vector $\mathbf{LMP} = (\mathrm{LMP}_i)_{i=1}^n \in \mathbb{R}^n$ can be expressed as (see Section 1.3.2, Definition 1.1 and Eq. (1.33))

$$\mathbf{LMP} = \lambda_{\text{en}}\mathbf{1} + \widehat{\mathbf{V}}^\top \boldsymbol{\mu} \in \mathbb{R}^n, \tag{6.6}$$

where $\boldsymbol{\mu} = \boldsymbol{\mu}^- - \boldsymbol{\mu}^+ \in \mathbb{R}^m$. Eq. (6.6) allows us to make the following remarks. First, note that $\mu_\ell = 0$ if and only if line ℓ is not congested, that is, if and only if $\underline{f}_\ell < f_\ell < \bar{f}_\ell$. In particular, $\mu_\ell^+ > 0$ if $f_\ell = \bar{f}_\ell$, and $\mu_\ell^- < 0$ if $f_\ell = \underline{f}_\ell$.[1] As a consequence, if there are no congested lines, the LMPs at all nodes are equal, i.e.,

$$\mathrm{LMP}_i = \lambda_{\text{en}}\ \forall i = 1\ \ldots, n,$$

[1] Note that μ^-, μ^+ cannot be both strictly positive, since lower and upper line flow constraints cannot be simultaneously binding.

and the common value λ_{en} in (6.6) is known as the marginal *energy component*. The energy component λ_{en} reflects the marginal cost of energy at the reference bus [101]. On the other hand, if some lines are congested, the LMPs will in general be different at different nodes (see Fig. 6.1), and the term $\tilde{\boldsymbol{\pi}} := \widehat{\mathbf{V}}^\top \boldsymbol{\mu}$ in Eq. (6.6) is called the marginal *congestion component*.

When ISOs calculate LMPs, they also include a *loss component*, which is related to the heat dissipated on transmission lines and is not accounted for by the DC-OPF model. The loss component is typically negligible compared to the other price components [172], and its inclusion goes beyond the scope of this chapter.

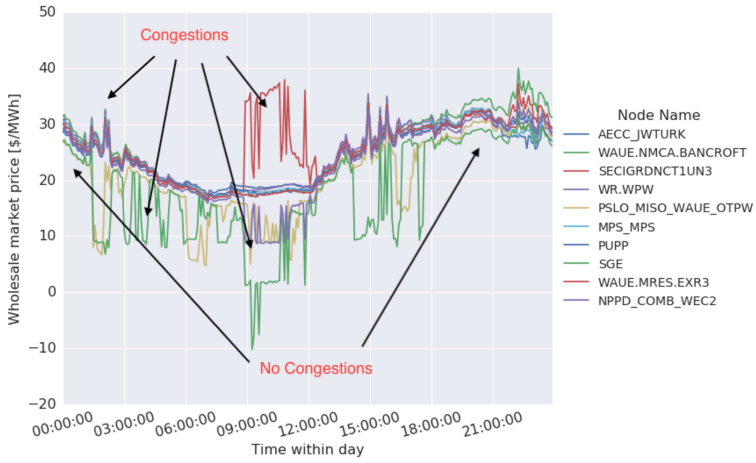


Figure 6.1: RT LMPs for 10 randomly selected nodes in the Southern Power Pool market [154].

The **LMP** vector also implicitly depends on the dual variables $\boldsymbol{\tau}^-, \boldsymbol{\tau}^+$ for generation constraints, which determine the *marginal* status of generators. A generator i is called *marginal* if $\underline{g}_i < g_i < \bar{g}_i$, in which case $\boldsymbol{\tau}_i^- = \boldsymbol{\tau}_i^+ = 0$, and *saturated* otherwise. The reason for this terminology is that the marginal cost of a marginal generator determines the LMP at its node. Indeed, denoting by $\mathcal{L}$ the Lagrangian function of the DC-OPF defined in (1.29), the stationarity condition for the optimal solution of the OPF reads

$$\frac{\partial \mathcal{L}}{\partial \mathbf{g}}\Big|_{\mathbf{g}^*} = \frac{\partial \mathbf{J}}{\partial \mathbf{g}}\Big|_{\mathbf{g}^*} - \lambda_{\text{en}} - \widehat{\mathbf{V}}^\top \boldsymbol{\mu} - \boldsymbol{\tau}^- + \boldsymbol{\tau}^+ = \mathbf{0}, \tag{6.7}$$

showing that $\text{LMP}_i = \lambda_{\text{en}} + (\widehat{\mathbf{V}})_i^\top \boldsymbol{\mu} = J_i'(g_i^*)$ is equal to the marginal cost of generation at node i if $\tau_i^- = \tau_i^+ = 0$n.

6.1.3 Problem statement

The goal of this chapter is to derive a structured machine learning methodology to predict LMPs by using only publicly available data. In other words, we assume we have no knowledge of the power system parameters that appear in the OPF formulation in Eqs. (6.1) - (6.4), such as grid topology, objective functions, line limits etc. The rationale for this choice lies in the observation that market participants usually do not have access to such proprietary data. However, the topic of decentralized price forecasts from a market participant perspective is becoming increasingly relevant due the development of Distributed Energy Resource technologies, which enable the owners of controllable energy assets to shape their wholesale market participation responsively and in a coordinated manner [59, 145]. To address the environmental and operational challenges, an important question concerns whether wholesale market prices could be inferred from the supply/demand mix on the grid and, then, used to create a feedback for "shaping" energy asset's production or consumption.

We take a *functional* viewpoint, i.e., we view the demand and renewable generation supply as variable parameters $\boldsymbol{\theta}$, or input, of the OPF. In particular, we consider a setting where the objective function, PTDF matrix, line limits and generation constraints are assumed to be fixed, although *unknown*, parameters of the OPF. Conversely, the demand and renewable generation supply correspond to a *variable parameter* $\boldsymbol{\theta}$, upon which the solution of the OPF and the $\mathbf{LMP}$ vector depend. The structure of our parametric approach is described in more details in Section 6.2.

The collection $\{\boldsymbol{\theta}(t), \mathbf{LMP}(t)\}_{t\in\mathcal{T}}$, where $\mathcal{T}$ is a finite discrete set $\mathcal{T} = \{1, \dots, T\}$, constitutes the historical dataset that we use to train and validate the machine learning methodology. In the rest of this chapter we focus on *hourly* LMPs prediction, therefore interpreting each time index t as a 60-minute interval. In particular, we assume $T = 24 \times n_{\text{days}}$, $n_{\text{days}} \in \mathbb{N}$. We observe, however, that the methodology developed in this chapter can be applied to different time windows as well, since the supply-demand matching process used to calculate the prices has the same structure across different time granularities. For the same reason, the approach is valid both for DA and RT price prediction.

6.1.4 Contribution of this chapter

In Section 1.4.3 we have reviewed the relevant literature on LMP forecasting, with a focus on the difference between the centralized and the decentralized perspectives. With that discussion in mind, we now summarize the main contributions of this chapter:

- By taking the market participants' point of view, we propose a decentralized price forecasting methodology that utilizes only publicly available data.

- By combining the theoretical insights on how system operators derive LMPs, and the latest developments in machine learning and convex optimization literature, we develop a novel algorithm that can predict market prices with high accuracy.

- The new approach reveals interesting and potentially very useful insights about different grid state regimes (both in terms of the grid wide generation/load mix, as well as line congestions) and their impact on prices across all nodes in the network. However, we observe in the extended paper [154] that the methodology has some limitations when it comes to predicting *price spikes*.

6.2 Multiparametric programming background

As discussed in Section 6.1.3, locational marginal prices can be thought of as deterministic functions of certain parameters $\boldsymbol{\theta}$, such as nodal demand and renewable generation supply. In this section we formalize this notion using the language of Multiparametric Programming Theory (MPT) [11], which, together with an additional assumption (see Section 6.3, and Assumption 6.1 in particular), allows us to parametrize LMPs by a vector of publicly available grid-level renewable supply and load.

Multiparametric Programming Theory is concerned with the study of optimization problems which depend on a vector of parameters, and aims at analyzing the impact of such parameters on the outcome of the problem, both in terms of primal and dual solutions. In our setting, the vector of parameters reads $\boldsymbol{\theta} = [\mathbf{d}^\top, \overline{\mathbf{w}}^\top]^\top \in \mathbb{R}^{n_d+n_w}$, where $\mathbf{d} \in \mathbb{R}^{n_d}$ is the demand vector, and $\overline{\mathbf{w}} \in \mathbb{R}^{n_w}$ is the vector of renewable generation available supply, i.e., $\overline{\mathbf{w}} = (\bar{g}_i)_{i\in\mathcal{N}_w}$ for $i \in \mathcal{N}_w$.

In this setting, the OPF problem in Eqs. (6.1) - (6.4), henceforth referred to as OPF($\boldsymbol{\theta}$), can be formulated as a standard Multiparametric Quadratic Program (MPQ) as follows:

$$\min_{\mathbf{g}\in\mathbb{R}^n} \quad \frac{1}{2}\mathbf{g}^\top\mathbf{H}^\top\mathbf{g} + \mathbf{g}^\top\mathbf{h} \tag{6.8}$$

$$\text{s.t.} \quad \mathbf{A}\mathbf{g} \leq \mathbf{b} + \mathbf{E}\boldsymbol{\theta}, \tag{6.9}$$

where $\mathbf{A} \in \mathbb{R}^{(2+2m+2n)\times n}, \mathbf{E} \in \mathbb{R}^{(2+2m+2n)\times(n_d+n_w)}, \mathbf{b} \in \mathbb{R}^{2+2m+2n}$, are suitably defined matrices and vectors, given below:

$$\mathbf{A} = \left[\begin{array}{cc} \mathbf{1}^\top & \\ -\mathbf{1}^\top & \\ \widehat{\mathbf{V}} & \\ -\widehat{\mathbf{V}} & \\ \hline I_{n_g} & \mathbf{0} \\ \mathbf{0} & I_{n_w} \\ \hline -I_{n_g} & \mathbf{0} \\ \mathbf{0} & -I_{n_w} \end{array}\right], \mathbf{b} = \left[\begin{array}{c} \mathbf{0} \\ \mathbf{0} \\ \bar{\mathbf{f}} \\ -\underline{\mathbf{f}} \\ \bar{\mathbf{g}}_{\mathcal{N}_g} \\ \mathbf{0} \\ -\underline{\mathbf{g}}_{\mathcal{N}_g} \\ \mathbf{0} \end{array}\right], \mathbf{E} = \left[\begin{array}{cc} -\mathbf{1}^\top & \mathbf{0} \\ \mathbf{1}^\top & \mathbf{0} \\ -\widehat{\mathbf{V}}_{\mathcal{N}_d} & \mathbf{0} \\ \widehat{\mathbf{V}}_{\mathcal{N}_d} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I_{n_w} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{array}\right], \boldsymbol{\theta} = \left[\begin{array}{c} \mathbf{d} \\ \hline \overline{\mathbf{w}} \end{array}\right]. \tag{6.10}$$

Here, we denote by $\mathbf{I}_k \in \mathbb{R}^{k\times k}$ the identity matrix of dimension k, by $\mathbf{0}$ a zero matrix of appropriate dimension, and we define $\bar{\mathbf{g}}_{\mathcal{N}_g} = (\bar{g})_{i\in\mathcal{N}_g}, \underline{\mathbf{g}}_{\mathcal{N}_g} = (\underline{g})_{i\in\mathcal{N}_g}$. Moreover, $\widehat{\mathbf{V}}_{\mathcal{N}_d} \in \mathbb{R}^{m\times n_d}$ denotes the submatrix of $\widehat{\mathbf{V}}$ obtained by selecting only the columns corresponding to nodes in $\mathcal{N}_d$.

A key result in MPT [181] is that the feasible parameter space $\boldsymbol{\Theta} \subseteq \mathbb{R}^{n_d+n_w}$ of the problem Eqs. (6.8) - (6.9) can be partitioned into a finite number of convex polytopes, each corresponding to a different *system pattern* [209]. A system pattern consists of a grid-wide state vector that indicate the saturated status of generators and congestion status of transmission lines. In order to state the result, we first introduce some definitions. System patterns can be formalized using the MPT concept of *optimal partitions*.

Definition 6.1 (Optimal Partition)**.** *Given a parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, let $\mathbf{g}^* = \mathbf{g}^*(\boldsymbol{\theta})$ denote the optimal generation vector obtained by solving the problem defined by Eqs.* (6.8) - (6.9)*. Let $\mathcal{J}$ denote the index set of constraints in Eq.* (6.9)*, with $|\mathcal{J}| = 2 + 2m + 2n$. The optimal partition of $\mathcal{J}$ associated with $\boldsymbol{\theta}$ is the partition $\mathcal{J} = \mathcal{B}(\boldsymbol{\theta}) \sqcup \mathcal{B}^{\complement}(\boldsymbol{\theta})$, with*

$$\mathcal{B}(\boldsymbol{\theta}) = \{1\}\{i \in \mathcal{J} \mid \mathbf{A}_i\mathbf{g}^* = \mathbf{b} + \mathbf{E}_i\boldsymbol{\theta}\}, \tag{6.11}$$

$$\mathcal{B}^{\complement}(\boldsymbol{\theta}) = \{i \in \mathcal{J} \mid \mathbf{A}_i\mathbf{g}^* < \mathbf{b} + \mathbf{E}_i\boldsymbol{\theta}\}. \tag{6.12}$$

The sets $\mathcal{B}$ and $\mathcal{B}^{\complement}$, respectively, correspond to binding and non-binding constraints of the OPF. With a minor abuse of notation, we identify the optimal partition $(\mathcal{B}, \mathcal{B}^{\complement})$ with the corresponding set of binding constraints $\mathcal{B}$. Given an optimal partition $\mathcal{B}$, let $\mathbf{A}_\mathcal{B}, \mathbf{E}_\mathcal{B}$ denote the submatrices of $\mathbf{A}$ and $\mathbf{E}$ containing the rows $\mathbf{A}_i, \mathbf{E}_i$ indexed by $i \in \mathcal{B}$, respectively.

Remark 6.1. *The energy balance equality constraint* (6.2) *in the original OPF formulation is rewritten as two inequalities indexed by $i = 1, 2$ in* (6.9)*, which are*

always binding, i.e., they read $\mathbf{A}_i\mathbf{g}^* = \mathbf{b}_i + \mathbf{E}_i\boldsymbol{\theta}$, $i = 1, 2$. *Looking at Eq.* (6.10), *we see that the two equations* $\mathbf{A}_i\mathbf{g}^* = \mathbf{b}_i + \mathbf{E}_i\boldsymbol{\theta},\ i = 1, 2,$ *are identical, and thus one of them is redundant. In the rest of this chapter, we eliminate one of the redundant constraints from the set* $\mathcal{B}$, *namely the one corresponding to* $i = 2$. *Therefore, we write* $\mathcal{B} = \{1\} \sqcup \mathcal{B}^{(cong)} \sqcup \mathcal{B}^{(sat)}$, *where* $\mathcal{B}^{(cong)} \subseteq \{3, \ldots, 2 + 2m\}$ *describes the congestion status of transmission lines, and* $\mathcal{B}^{(sat)} \subseteq \{2 + 2m + 1, 2 + 2m + 2n\}$ *describe the saturated status of generators.*

Definition 6.2. *Given an optimal partition* $\mathcal{B}$, *we say that the linear independent constraint qualification (LICQ) holds if the set of active constraints gradient are linearly independent, i.e., if* $\mathbf{A}_{\mathcal{B}} \in \mathbb{R}^{|\mathcal{B}| \times n}$ *has full row rank.*

The following result, originally established within the multiparametric programming literature (see [181], Theorem 1), and subsequently stated in [209] in the context of LMP forecasting, constitutes the theoretical foundation of the methodology presented in this chapter.

Theorem 6.1. *Assume that* $\mathbf{H}$ *is positive definite,* $\boldsymbol{\Theta}$ *is a full dimensional compact set, and that the LICQ regularity condition is satisfied for every* $\boldsymbol{\theta}$ *in* $\boldsymbol{\Theta}$. *Then,* $\boldsymbol{\Theta}$ *can be covered by a union of a finite number* M *of full-dimensional compact convex polytopes* $\boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_M$, *referred to as critical regions, such that:*

- *their interiors are pairwise disjoint*

$$\mathring{\boldsymbol{\Theta}}_k \cap \mathring{\boldsymbol{\Theta}}_h = \emptyset \quad , \forall k \neq h,$$

  and each interior $\mathring{\boldsymbol{\Theta}}_k$ *corresponds to the largest set of parameters yielding the same optimal partition.*

- *within the interior of each critical region* $\mathring{\boldsymbol{\Theta}}_k$, *the optimal generation* $\mathbf{g}^*$ *and the associated* $\mathbf{LMP}$ *vector are affine functions of* $\boldsymbol{\theta}$.

- *the map* $\boldsymbol{\Theta} \ni \boldsymbol{\theta} \rightarrow \mathbf{LMP}(\boldsymbol{\theta})$ *defined over the entire parameter space is piecewise affine and continuous.*

6.3 Overview of the prediction methodology

The theoretical framework described in Section 6.2 implicitly assumes a centralized viewpoint, since the calculation of the critical regions $\boldsymbol{\Theta}_k$ and of the corresponding affine functions $\boldsymbol{\Theta}_k \ni \boldsymbol{\theta} \rightarrow \mathbf{LMP}$ requires the knowledge of all the fixed parameters of the OPF, such as the grid topology (encoded in the matrix $\widehat{\mathbf{V}}$), generators' bids (encoded in $\mathbf{H}, \mathbf{h}, \bar{\mathbf{g}}, \underline{\mathbf{g}}$), and line limits $\bar{\mathbf{f}}, \underline{\mathbf{f}}$, as well

as access to historical data for $\boldsymbol{\theta}$, which consists of nodal demand and available renewable supply.

From a decentralized perspective, however, the aforementioned system parameters are not available. Even if they were, and thus one could characterize the maps $\boldsymbol{\Theta}_k \ni \boldsymbol{\theta} \rightarrow \mathbf{LMP}$ analytically, it would not be clear how to use them, since accurate historical data for nodal demand and nodal renewable supply (which constitute the vector $\boldsymbol{\theta}$) are also unavailable.

Our novel prediction methodology deals with both of the aforementioned problems by assuming no information on system parameters, and only relying on aggregated grid-level data that are publicly accessible. The publicly available market data depends on the specific market and commonly includes historical aggregated grid-level load, aggregated grid-level supply from various energy sources (i.e., wind, solar, coal etc.), and nodal LMPs. Specifically, in the rest of this chapter, we assume that the available data are the grid-wide demand $D(t) = \sum_{i \in \mathcal{N}_d} d_i(t)$, grid-wide available renewable supply $W(t) = \sum_{i=1\ldots,n_w} \overline{w}_i(t)$, and historical LMPs. The developed methodology requires at least the aforementioned components. More granular data (e.g. nodal load and generation) would improve the accuracy of the algorithm, but are not essential.

In order to use the MPT framework described in Section 6.2, we need to relate the nodal demand and renewable supply described by the parameter $\boldsymbol{\theta}(t) \in \mathbb{R}^{n_d+n_w}$, for $t = 1, \ldots, T$, to the corresponding grid-level quantities $\mathcal{M}(t) := [D(t), W(t)]^\top \in \mathbb{R}^2$, which will be referred to as $\mathcal{M}$ *vectors* from now on. We do that by (i) introducing the concept of $\mathcal{M}$*-regime*, and (ii) assuming a specific relationship between nodal and global quantities.

A $\mathcal{M}$-regime consists of the set of input parameters corresponding to the same hour of the day, namely $\mathcal{R}_h^{(\text{mix})} = \{\boldsymbol{\theta}(t)\}_{t \in \mathcal{T}(h)}$, and $\mathcal{T}(h) = \{h, 24 + h, \ldots, 24(n_{\text{days}} - 1) + h\}$, for $h = 1, \ldots, 24$. The rationale behind this definition is that demand and renewable supply for the same hour of the day exhibit significant correlation across different days. For example, solar generation is always zero during night hours, while total load usually peaks in the evening around the same time.

The concept of $\mathcal{M}$-regime, in its essence, tries to capture the fact that historical demand and supply that share similarities can be grouped in clusters, and doing so can improve the performance of the algorithm. For ease of exposition, in this chapter we present a very simple clustering rule based on the hour of the day, and refer to the extended paper [154] for possible generalizations. Next, we introduce the following assumption.

Assumption 6.1. *Within each $\mathcal{M}$-regime $\mathcal{R}_h^{(mix)}$, all nodal loads preserve the same consumption ratio with respect to the total load, namely $d_i(t) = \alpha_i^{(d)} D(t)$, $\forall i \in$*

$\mathcal{N}_d$, $\forall t \in \mathcal{T}(h)$, *with* $0 \leq \alpha_i^{(d)} \leq 1$, $\sum_{i \in \mathcal{N}_d} \alpha_i^{(d)} = 1$. *Similarly, we assume that* $\overline{w}_i(t) = \alpha_i^{(w)} W(t)$, $\forall j = 1, \ldots, n_w$, *with* $0 \leq \alpha_j^{(w)} \leq 1$, $\sum_{j=1,\ldots,n_w} \alpha_j^{(w)}$, $\forall t \in \mathcal{T}(h)$.

Intuitively, Assumption 6.1 states that, within the same intra-day $\mathcal{M}$-regime, each renewable generator preserves constant production fraction in relation to the total renewable supply, and similarly for demand. For example, when wind generation increases during a specific hour, we assume that all wind generators produce proportionally more power.

This simplifying assumption enables us to extend the piecewise affinity established in Theorem 6.1 to our decentralized framework and parametrize LMPs using $\mathcal{M}$ vectors. More specifically, Theorem 6.1 establishes that nodal LMPs are piecewise affine functions of the nodal quantity $\boldsymbol{\theta} = [\mathbf{d}^\top, \overline{\mathbf{w}}^\top]^\top$, while Assumption 6.1 allows to define a one-to-one mapping between $\boldsymbol{\theta}$ and the grid-level quantity $\mathcal{M} = [D, W]^\top$, yielding that the $\mathbf{LMP}$ vector is also a piecewise affine function of the $\mathcal{M}$ vectors within each $\mathcal{M}$-regime. We observe that, by requiring Assumption 6.1 to hold in each individual $\mathcal{M}$-regime separately, rather than globally, we make it less restrictive. This is indeed one of the reasons to define $\mathcal{M}$-regimes in the first place, together with the other advantages discussed later in Section 6.4.

In order to learn the function $\mathcal{M} \to \mathbf{LMP}(\mathcal{M})$ we conveniently utilize a statistical procedure for fitting adaptive regression splines, called Multivariate Adaptive Regression Splines (MARS) [66]. Relying on the piecewise affinity of the LMP functions, MARS identifies a linear combination of truncated spline functions of the form $\max\{x_j - q, 0\}$ and $\max\{q - x_j, 0\}$, where $x_j \in \mathbb{R}$ are the relevant covariates (in our case, grid-level demand and renewable supply), while the q's are knot locations (critical region switching points) identified by the algorithm.

Provided with enough "uniform" data, and under Assumption 6.1, the MARS algorithm would accurately recover the piecewise affine function $\mathcal{M} \to \mathbf{LMP}(\mathcal{M})$. In this context, by "uniform" data we mean that samples of $\mathcal{M}$ vectors from each critical region are available in approximately equal proportions. One factor that can negatively affect the performance of MARS is indeed the scarcity of data corresponding to specific regions. If a critical region is "rare", there will not be enough historical datapoints to correctly learn the affine function within that region.

Moreover, the higher the number of pieces defining the overall map $\mathcal{M} \to \mathbf{LMP}(\mathcal{M})$, the harder it is for the algorithm to achieve good performance without increasing the size of the dataset. Conversely, MARS usually performs better when learning a smaller number of pieces. For example, if one would know beforehand that a certain subset of data corresponds to a given critical

region $\mathbf{\Theta}_i$, then on this specific subset the problem reduces to the much easier task of learning a single affine function $\mathbf{\Theta}_i \ni \mathcal{M} \to \mathbf{LMP}(\mathcal{M})$.

Unfortunately, the exact computation of critical regions requires full information on system parameters, thus preventing us from clustering $\mathcal{M}$ vectors based on the critical region they belong to. However, it turns out that we can use historical nodal prices to *infer* the set of congested lines, henceforth referred to as *congestion regimes*, corresponding to specific subsets of data. More specifically, given a set of congested lines $\mathcal{B}^{(\text{cong},*)}$ (see Remark 6.1), we define the associated congestion regime as the collection of corresponding time instances, namely

$$\mathcal{R}^{(\text{cong})} := \{t \in \mathcal{T} \,:\, \mathcal{B}^{(\text{cong})}(\boldsymbol{\theta}(t)) = \mathcal{B}^{(\text{cong},*)}\}. \tag{6.13}$$

In Section 6.4 we describe how we can recover the congestion regimes $\mathcal{R}^{(\text{cong})}$ by only looking at historical prices. As a result, after grouping the data in $\mathcal{M}$-regimes, we are able to further cluster them according to the corresponding congestion regime. This, in turn, allows us to significantly reduce the number of pieces that MARS has to learn, since within a fixed congestion regime the only remaining pieces are those describing the status of saturated generator $\mathcal{B}^{(\text{sat})}$.

To summarize, the congestion recovery and the MARS algorithm constitute the two main components of our methodology. In the next two sections, we examine each component in more detail.

6.4 Congestion regimes recovery

Based on the definition of the PTDF matrix in Eq. (1.18) and of the LMP vector in (6.6), the marginal congestion price vector (excluding the reference bus, for which the congestion component is always equal to 0) at a given time t can be represented as

$$\boldsymbol{\pi}(t) = \widetilde{\mathbf{L}}^{-1}\widetilde{\mathbf{A}}^{\top}\mathbf{D}\boldsymbol{\mu}(t) = \widetilde{\mathbf{L}}^{-1}\mathbf{s}(t) \in \mathbb{R}^{n-1}, \tag{6.14}$$

with $\mathbf{s}(t) = \widetilde{\mathbf{A}}^{T}\mathbf{D}\boldsymbol{\mu}(t)$ and $\boldsymbol{\mu}(t) = \boldsymbol{\mu}^{-}(t) - \boldsymbol{\mu}^{+}(t)$. The vector $\mathbf{s}(t) \in \mathbb{R}^{n-1}$ contains the information on the congested lines, since

$$\mathbf{s}(t) = \sum_{\ell=1}^{m} \mu_\ell(t) x_\ell^{-1} \tilde{\mathbf{a}}_\ell, \tag{6.15}$$

where $\tilde{\mathbf{a}}_\ell \in \mathbb{R}^{n-1}$ is the ℓ-th column of $\widetilde{\mathbf{A}}^{T}$. The non-zero entries of $\mathbf{s}(t)$ represent nodes corresponding to congested transmission lines. Thus, by stacking historical $\boldsymbol{\pi}(t), \mathbf{s}(t)$ for T different time intervals as columns of the matrices $\mathbf{\Pi}, \mathbf{S} \in \mathbb{R}^{(n-1)\times T}$, we can rewrite ((6.14)) in matrix form as

$$\mathbf{\Pi} = \widetilde{\mathbf{L}}^{-1}\mathbf{S}. \tag{6.16}$$

In the following, we use the previous relationship and the properties of matrices $\widetilde{\mathbf{L}}$ and $\mathbf{S}$, henceforth referred to as *topology matrix* and *congestion matrix*, respectively, to recover diverse congestion regimes that occur in the grid.

The rest of this section is based on the work in [101], where the authors derive a methodology to recover matrices $\widetilde{\mathbf{L}}$ and $\mathbf{S}$ via the price matrix $\mathbf{\Pi}$ and the relation (6.16). We point out that in [101] the goal is to infer an unknown power grid topology by leveraging only publicly available data, and thus the authors focus specifically on the topology matrix $\widetilde{\mathbf{L}}$. On the other hand, in our setting, we are mostly interested in inferring the congestion status of transmission lines, rather than the entire grid topology, hence we shift the attention to the congestion matrix $\mathbf{S}$. The underlying methodology is the same for both tasks, and is summarized below for the sake of completeness of this chapter.

Note that the problem of recovering matrices $\widetilde{\mathbf{L}}$ and $\mathbf{S}$ by only knowing the price matrix $\mathbf{\Pi}$ in Eq. (6.16), is clearly under-determined. The way to approach this challenge is to exploit *structural* properties of the matrices, which we now describe. First, observe that matrix $\widetilde{\mathbf{A}} \in \mathbb{R}^{m\times(n-1)}$ is a full column-rank matrix, and $\widetilde{\mathbf{L}} \in \mathbb{R}^{(n-1)\times(n-1)}$ is strictly positive definite with non-positive off-diagonal entries.

Second, note that matrices $\widetilde{\mathbf{L}}$ and $\mathbf{S}$ satisfy the following structural properties: (i) $\widetilde{\mathbf{L}}$ is a positive definite M-matrix[2] and is sparse, and (ii) $\mathbf{S}$ is sparse and low-rank. The sparsity of $\widetilde{\mathbf{L}}$ follows from the fact that the graph underlying a power grid is usually sparingly connected, especially for US power grids [5]. The fact that $\mathbf{S}$ is sparse and low-rank follows from Eq. (6.15) and the fact that usually only a very small subset of transmission lines get congested [101], implying that most of the terms in the sum in (6.15) are zero. In [101], the authors suggest to recover matrices $\widetilde{\mathbf{L}}$ and $\mathbf{S}$ by solving the optimization problem:

$$\begin{aligned} \min_{\widetilde{\mathbf{L}},\mathbf{S}} \quad & \|\mathbf{S}\|_0 + \kappa_0 \|\widetilde{\mathbf{L}}\|_0 \\ \text{s.t.} \quad & \widetilde{\mathbf{L}}\mathbf{\Pi} = \mathbf{S},\ \widetilde{\mathbf{L}} \succ \mathbf{0}, \widetilde{\mathbf{L}} \leq \mathbf{I}, \end{aligned} \tag{6.17}$$

where $\|\mathbf{X}\|_0$ is the ℓ_0 pseudo-norm counting the non-zero entries of matrix $\mathbf{X}$, and $\kappa_0 \geq 0$. In words, the optimization problem (6.17) tries to find the *sparsest pair* $(\widetilde{\mathbf{L}}, \mathbf{S})$ that satisfies the structural properties mentioned before. Since this problem is in general NP-hard, it is suggested in [101] that the following convex relaxation be used:

[2] See Section 1 in [149] for the definition of M-matrix.

$$\begin{aligned} \min_{\widetilde{\mathbf{L}},\mathbf{S}} \quad & \|\mathbf{S}\|_1 + \kappa_1 \mathrm{tr}(\mathbf{P}\widetilde{\mathbf{L}}) - \kappa_2 \log|\widetilde{\mathbf{L}}| \\ \text{s.t.} \quad & \widetilde{\mathbf{L}}\mathbf{\Pi} = \mathbf{S},\ \widetilde{\mathbf{L}} \in \mathcal{C}, \end{aligned} \tag{6.18}$$

with $\|\mathbf{X}\|_1 = \sum_{i,j} |\mathbf{X}_{i,j}|$ denoting the ℓ_1 norm of matrix $\mathbf{X}$, $\mathbf{P} = \mathbf{I} - \mathbf{1}\mathbf{1}^T$, $\mathcal{C} := \{\widetilde{\mathbf{L}} : \widetilde{\mathbf{L}} \succeq \mathbf{0},\ \widetilde{\mathbf{L}} \leq \mathbf{I}\}$, $\kappa_1, \kappa_2 \geq 0$, and $\widetilde{\mathbf{L}} \succeq \mathbf{0}$ denoting a positive semidefinite matrix. For small grids, such as the IEEE 30-bus test case in Section 6.6, the previous semidefinite program can be solved quite efficiently using standard software libraries. In particular, we used CVX, a Matlab package for specifying and solving convex programs [76].

6.4.1 Recovery of matrix S in a given $\mathcal{M}$-regime

Besides the topology matrix $\widetilde{\mathbf{L}}$, the algorithm discussed in Section 6.4 allows us to recover the congestion matrix $\mathbf{S}$ as well. It turns out that it is beneficial to solve the optimization problem (6.18) within each $\mathcal{M}$-regime separately, due to the fact that $\mathcal{M}$ vectors corresponding to the same hour h are usually associated with a few different congestion regimes (see Fig. 6.2), implying that the matrix $\mathbf{S}(h)$, obtained from $\mathbf{S}$ by selecting columns corresponding to time indices in $\mathcal{T}(h)$, is particularly sparse.

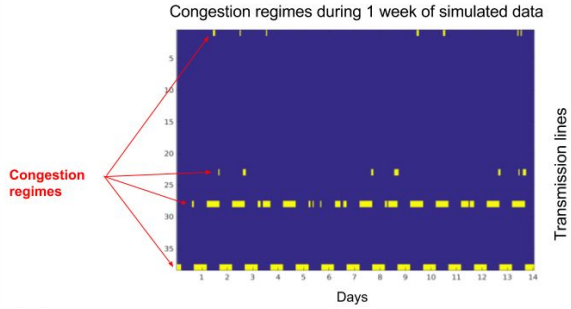
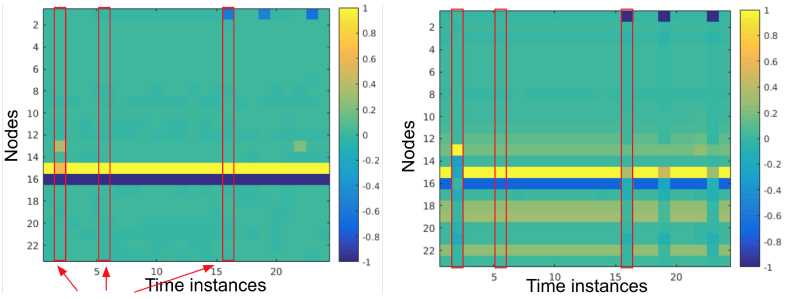


Figure 6.2: A segment of the congestion matrix $\mathbf{S}$, visualizing the congestion regimes during two weeks of simulated data for the IEEE 30-bus test case. The x-axis spans time instances at hourly granularity, while the y-axis corresponds to node indices. The non-zero entries (depicted in yellow) represent nodes connected to congested transmission lines, in view of Eq. (6.15). Zero entries are depicted in blue.

Fig. 6.3a depicts an example of a segment of the normalized version of matrix $\mathbf{S}(h)$, which is obtained by dividing each column j of $\mathbf{S}(h)$ by $\max_{i \in \mathcal{N}} S_{i,j}(h)$. The x-axis spans time instances, while the y-axis corresponds to node indices.

Most of the entries of the matrix $\mathbf{S}$ are equal to zero, with a few entries having relatively large absolute values. Those entries represent nodes connected to the congested transmission lines, in view of Eq. (6.15). As a consequence, the j-th column of $\mathbf{S}(h)$ uniquely identifies the set of congested lines at time $24 \cdot (j-1) + h \in \mathcal{T}(h)$.



(a) Segment of (normalized) congestion matrix $\mathbf{S}(h)$.

(b) Segment of (normalized) recovered congestion matrix $\widetilde{\mathbf{S}}(h)$.

Figure 6.3: Comparison between actual and recovered congestion matrices.

The focus of this section is on recovering the congestion matrix $\widetilde{\mathbf{S}}(h)$ within each $\mathcal{M}$-regime by solving the optimization problem in Eq. (6.18). For each $\mathcal{M}$-regime, we obtain the corresponding matrix $\widetilde{\mathbf{S}}(h)$ by solving the optimization problem in Eq. (6.18). An example of a segment of the recovered congestion matrix $\widetilde{\mathbf{S}}(h)$ is shown in Fig. 6.3b. We can see how $\widetilde{\mathbf{S}}(h)$ correctly identifies the non-zero entries of the actual congestion matrix $\mathbf{S}(h)$, but introduces some noise that masks zero entries.

The congestion matrix encodes information that can be used to improve the accuracy of the MARS algorithm, as discussed in Section 6.3. More specifically, in view of (6.14), we recover congestion regimes by clustering the columns of $\widetilde{\mathbf{S}}(h)$ using a combination of Principal Component Analysis [98] and k-means clustering [114]. The red boxes marked by the arrows in Fig. 6.3 highlight the three different congestion regimes occurring during a subset of $\mathcal{T}(h)$, showing that the algorithm is able to correctly recover them.

6.5 Learning the mapping between $\mathcal{M}$ vectors and LMP vectors

The next component of the methodology relies on Assumption 6.1 and Theorem 6.1, and focuses on learning the piecewise affine function between the

vectors $\mathcal{M}(t) = [D(T), W(t)]^\top$ and the **LMP** vector, as outlined in Section 6.3.

Mapping $\mathcal{M}$ vectors to LMP vectors within each congestion regime. First, we apply the MARS algorithm within each $\mathcal{M}$-regime and congestion regime separately. More specifically, given an $\mathcal{M}$-regime $\mathcal{R}_i^{(\text{mix})}$, let

$$\mathcal{F}(i) := \left\{\mathcal{R}_{i,1}^{(\text{cong})}, \ldots, \mathcal{R}_{i,n_i^{(\text{cong})}}^{(\text{cong})}\right\},$$

be the set of associated congestion regimes, $n_i^{(\text{cong})} \geq 1$. For each pair (i, j) such that $\mathcal{R}_{i,j}^{(\text{cong})} \in \mathcal{F}(i)$, we use the MARS algorithm to learn the piecewise affine maps

$$\phi_{i,j}^{(\text{lmp})} : \mathbb{R}^2 \ni \mathcal{M} \to \mathbf{LMP} \in \mathbb{R}^n, \tag{6.19}$$

where each piece is determined by a different saturation regime (see Section 6.3).

Mapping $\mathcal{M}$ vectors to congestion regimes within each $\mathcal{M}$-regime. During the prediction stage, we will use the learned functions $\phi_{i,j}^{(\text{lmp})}$ to map forecasted $\mathcal{M}$ vectors for hour i to the associated **LMP** vectors. In order to do that, we need to be able to assign the never-seen $\mathcal{M}$ vectors to the corresponding congestion regimes $\mathcal{R}^{(\text{cong})}$. To this end, given an $\mathcal{M}$-regime $\mathcal{R}_i^{(\text{mix})}$, we relate the corresponding $\mathcal{M}$-vectors to one of the congestion regimes in

$$\mathcal{F}(i) = \left\{\mathcal{R}_{i,1}^{(\text{cong})}, \ldots, \mathcal{R}_{i,n_i^{(\text{cong})}}^{(\text{cong})}\right\}$$

by training a classification algorithm. Thanks to the geometrical structure prescribed by Theorem 6.1 and the proportionality Assumption 6.1, the sets of $\mathcal{M}$ vectors corresponding to different congestion regimes can be accurately classified using, for example, the Support Vector Machine (SVM) algorithm, leading to the classification rule

$$\phi_i^{(\text{cong})} : \mathcal{R}_i^{(\text{mix})} \ni \mathcal{M} \to \mathcal{R}_{i,j}^{(\text{cong})} \in \mathcal{F}(i).$$

6.5.1 Summary of the methodology pipeline

Finally, we summarize the finalized methodology by splitting it into the training and prediction stages.

Training stage:

Step T.0: Collect historical $\mathcal{M}$ vectors $\mathcal{M}(t) = [D(t), W(t)]^\top$, $t \in \mathcal{T}$, as well as historical nodal **LMP** vectors, and cluster the $\mathcal{M}$ vectors in $\mathcal{M}$-regimes (see Section 6.3).

Step T1: Using historical price data, perform recovery of topology matrix $\widetilde{\mathbf{L}}$ and congestion matrix $\mathbf{S}(i)$ within each $\mathcal{M}$-regime i (see Section 6.4).

Step T2: For each $\mathcal{M}$-regime i, perform k-means clustering of the columns of the recovered congestion matrix $\tilde{\mathbf{S}}(i)$ to obtain the congestion regimes $\mathcal{F}(i) = \{\mathcal{R}_{i,1}^{(\text{cong})}, \ldots, \mathcal{R}_{i,n_i^{(\text{cong})}}^{(\text{cong})}\}$ (see Section 6.4.1).

Step T3: For each $\mathcal{M}$-regime i, learn the classification rule $\phi_i^{(\text{cong})}$ that maps $\mathcal{M}$-vectors to the associated congestion regime $\mathcal{R}_{i,j}^{(\text{cong})} \in \mathcal{F}(i)$ using the SVM algorithm (see Section 6.5).

Step T4: For each (i,j) such that $\mathcal{R}_{i,j}^{(\text{cong})} \in \mathcal{F}(i)$, use MARS to learn the piecewise affine function $\phi_{i,j}^{(\text{lmp})}$ mapping $\mathcal{M}$ vectors to **LMP** vectors (see Section 6.5).

Prediction stage:

Step P0: Obtain $\mathcal{M}$-vectors forecasts, and map them to the matching $\mathcal{M}$-regime i.

Step P1: Within each $\mathcal{M}$-regime i, use the trained classification model in step **T3** to assign $\mathcal{M}$ forecasts to the corresponding congestion regime $\mathcal{R}_{i,j}^{(\text{cong})} \in \mathcal{F}(i)$.

Step P2: For each (i,j) such that $\mathcal{R}_{i,j}^{(\text{cong})} \in \mathcal{F}(i)$, use the trained MARS models in step **T4** to map the $\mathcal{M}$ forecast to the resulting **LMP** vector forecast.

6.6 Numerical case study and validation

In this section we report validation results for the IEEE 30-bus test system of MATPOWER [210]. The test case includes all parameters needed to run DC-OPF, but it does not include renewable generators. For this reason, we modify a subset of the original generators (specifically, the generators located at nodes 2 and 27) as follows: (i) we set the corresponding cost functions equal to a very small positive number,[3] and ii) we let the corresponding generation limit $\bar{g}$ vary according to a distribution learned from CAISO data [32], as explained below.

Moreover, we uniformly multiply the transmission line limits by a small factor, set at 1.2, to ensure that we have few congestion events, in line with the empirical observation that congestion events are sparse [101]. Finally, we

[3] In order to model the fact that renewable generators are usually much cheaper than conventional ones.

scale up the generation limits for the other generators uniformly by a factor 3 to avoid scenarios where there is not enough total supply to match demand, which would lead to an unfeasible OPF.

6.6.1 Generation of historical $\mathcal{M}$ vectors

For demand profiles, we download 3 months of historical total demand profiles from S&P Global [173], collected at *hourly granularity*. We then fit a multivariate Gaussian distribution to the historical data and obtain a model $\mathbf{d}^{(\text{tot})} \sim N_{24}(\mu_{\text{dem}}, \Sigma_{\text{dem}})$, from which we can sample daily total load profiles. Next, each daily total load realization is scaled down so that its hourly average value matches the base-level demand of the test case. Finally, for each load node i, we set its demand to be equal to a fixed fraction $\alpha_i^{(d)}$ of the total demand, where $\sum_i \alpha_i^{(d)} = 1, 0 \leq \alpha_i^{(d)} \leq 1$.

To simulate renewable generation profiles, we download a daily profile for solar generation in the California ISO [32], which we denote as $\mathbf{w}^{(\text{tot},0)} \in \mathbb{R}^{24}$. A daily realization is then obtained by sampling from normal distribution

$$\mathbf{w}^{(\text{tot})} \sim \mathcal{N}_{24}(\mathbf{w}^{(\text{tot},0)}, \text{diag}(\mathbf{w}^{(\text{tot},0)})\sigma_w^2),$$

where we set $\sigma_w^2 = 0.1$. Next, the generated profiles are scaled down to be consistent with the base level of the other generators. Finally, the renewable generation capacity $\bar{g}_j$ for a specific renewable node j is set to be equal to a fixed fraction $\alpha_j^{(w)}$ of the corresponding grid-level quantity, where $\sum_j \alpha_j^{(w)} = 1$ and $0 \leq \alpha_j^{(w)} \leq 1$. The values of the fractions $\boldsymbol{\alpha}^{(d)}, \boldsymbol{\alpha}^{(w)}$ are based on the base-level data of the test case.

6.6.2 Generation of historical LMP vectors

In order to generate synthetic **LMP** vectors, which are needed for training and testing of our methodology, we run the DC-OPF in MATPOWER using the $\mathcal{M}$ vectors generated in the previous step. We recover the nodal quantities necessary to run the DC-OPF by using the proportionality Assumption 6.1 and the fractions $\boldsymbol{\alpha}^{(d)}, \boldsymbol{\alpha}^{(w)}$. The nodal quantities are only used to generate a synthetic *historical* dataset of **LMP** vectors,[4] and they play no further role in our validation.

[4] Which are publicly available in the case of actual wholesale energy market data.

6.6.3 Training and prediction stages

The training phase consists of steps **T1-T4** and we perform it using 3 months worth of synthetic $\mathcal{M}$ and **LMP** vectors. To evaluate the predictive performance of our approach, described in steps **P1-P3**, we generate synthetic day-ahead forecasts for $\mathcal{M}$ vectors as follows. First, we generate 100 samples from the multivariate Gaussian distributions for total demand and renewable supplies defined in Section 6.6.1, Next, in the prediction phase, we generate daily total demand $\mathbf{d}^{(\text{tot})}(i) \in \mathbb{R}^{24}$ and renewable supply $\mathbf{w}^{(\text{tot})}(i) \in \mathbb{R}^{24}$ over a testing period of 50 days, which we interpret as the *actual* realizations, and match each daily realization to the closest (in terms of the L2 norm) typical profile among the ones identified in the previous stage to obtain the *synthetic day-ahead forecasts* $\widehat{\mathbf{d}}^{(\text{tot})}(i)$ and $\widehat{\mathbf{w}}^{(\text{tot})}(i)$ for each day $i = 1, \ldots, 50$.

6.6.4 Performance evaluation

By stacking the 24-dimensional vectors $\mathbf{d}^{(\text{tot})}(i)$, $\widehat{\mathbf{d}}^{(\text{tot})}(i)$,$\mathbf{w}^{(\text{tot})}(i)$, $\widehat{\mathbf{w}}^{(\text{tot})}(i)$ for the different days in the testing period as columns of the matrices $\mathbf{D}^{(\text{tot})}$, $\widehat{\mathbf{D}}^{(\text{tot})}$, $\mathbf{W}^{(\text{tot})}$, $\widehat{\mathbf{W}}^{(\text{tot})} \in \mathbb{R}^{24,50}$, the forecasting relative errors are defined by

$$\text{err}^{(\text{dem})} = \frac{\|\mathbf{D}^{(\text{tot})} - \widehat{\mathbf{D}}^{(\text{tot})}\|_F}{\|\mathbf{D}^{(\text{tot})}\|_F}, \; \text{err}^{(\text{ren})} = \frac{\|\mathbf{W}^{(\text{tot})} - \widehat{\mathbf{W}}^{(\text{tot})}\|_F}{\|\mathbf{W}^{(\text{tot})}\|_F},$$

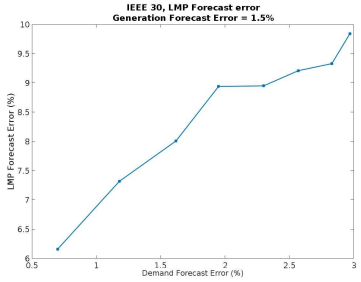
where $\|\mathbf{A}\|_F = \sqrt{\sum_{i,j} |a_{ij}|^2}$ denotes the Frobenius norm of matrix $\mathbf{A}$, which is equivalent to the ℓ_2 norm of the vector obtained by stacking the columns of the matrix $\mathbf{A}$ one under the other. Similarly, for each node $k = 1, \ldots, n$, we stack the actual and predicted LMPs as columns of the matrices $\mathbf{M}_k^{\text{lmp}}$, $\widehat{\mathbf{M}}_k^{\text{lmp}} \in \mathbb{R}^{24,50}$, and we evaluate the predictive performance of our methodology using the mean relative error across all nodes,defined as

$$\text{err}^{(\text{lmp})} := \frac{1}{n} \sum_{k=1}^{n} \text{err}_k^{(\text{lmp})}, \quad \text{err}_k^{(\text{lmp})} = \frac{\|\mathbf{M}_k^{\text{lmp}} - \widehat{\mathbf{M}}_k^{\text{lmp}}\|_F}{\|\mathbf{M}_k^{\text{lmp}}\|_F}.$$

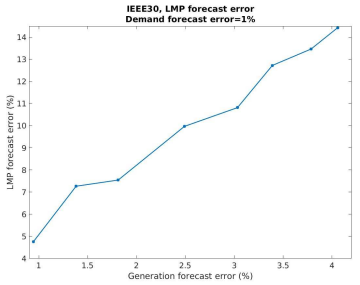
Figs. 6.4a and 6.4b capture the sensitivity of the predictive performance as a function of the forecasting errors in total load and renewable generation, showing that the algorithm is able to forecast future LMPs with reasonable accuracy. [5]

[5] For a comparison of the presented methodology with competing algorithms across a wider range of accuracy criteria, we refer to the paper [154], on which this chapter is based on. In particular, the paper [154] includes a validation on real market data from the Southwest Power Pool market.

More specifically, Fig. 6.4a (respectively, Fig. 6.4b (b)) shows how the prediction error changes as a function of the forecasting error in the total load (respectively, total renewable generation) where the generation (respectively, load) forecasting error is kept fixed.



(a) LMP prediction error $\text{err}^{(\text{lmp})}$ as a function of $\text{err}^{(\text{dem})}$, with $\text{err}^{(\text{ren})}$ fixed at 1.5%.



(b) LMP prediction error $\text{err}^{(\text{lmp})}$ as a function of $\text{err}^{(\text{ren})}$, with $\text{err}^{(\text{dem})}$ fixed at 1%.

6.7 Concluding remarks

In this chapter, we show that the wholesale energy prices can be inferred using limited, publicly available, historical market data. By utilizing the basic underlying physical model that captures generation-load matching on the grid, we develop a methodology for predicting locational marginal prices from a decentralized perspective, which also sheds light on the connection between market data and power grid congestion status. The methodology is validated on the IEEE 30-bus test case. While performing well in predicting intra-day variations, the proposed methodology may fail to predict rare occurrences of price spikes [154]. This is imputable to the scarcity of historical data corresponding to such rare events, which impacts the accuracy of the machine learning algorithms described in this chapter, and to the limitations of the simplifying Assumption 6.1. The issue of rare price spikes in LMPs is studied from a centralized perspective in Chapter 7.

Chapter 7

Large Deviations in Locational Marginal Prices

Contents

The problem of predicting energy price spikes, as opposed to expected intra-day variations, is a particularly difficult one, as discussed in Section 1.4.3. In this chapter, we investigate large fluctuations of Locational Marginal Prices (LMPs, see Definition 1.1) in wholesale energy markets caused by volatile renewable generation profiles. Specifically, we take the centralized perspective of the grid operator to study the probability of events of the form

$$\mathbb{P}\big(\mathbf{LMP} \notin \prod_{i=1}^{n} [\alpha_i^-, \alpha_i^+]\big), \tag{7.1}$$

where $\mathbf{LMP}$ is the vector of of LMPs at the n power grid nodes, and $\boldsymbol{\alpha}^-, \boldsymbol{\alpha}^+ \in \mathbb{R}^n$ are vectors of price thresholds specifying undesirable price occurrences.

We propose a novel approach combining multiparametric programming [181] with large deviations theory [55]. By exploiting the structure of the OPF problem, and assuming a centralized perspective, we first derive the deterministic piecewise affine, possibly discontinuous function linking the stochastic input process, modeling uncontrollable renewable generation, to the LMPs. This, in turn, allows us to utilize large deviations theory to identify the most likely ways for extreme price spikes to happen as a result of fluctuations of renewable generation.

The large deviations approach offers a powerful and flexible framework that holistically combines the network structure and operation paradigm with a stochastic model for renewable generation. This approach enables us to: i) approximate the probability of price spikes by means of solving a deterministic convex optimization problem, ii) rank the nodes of the power grids according to their likelihood of experiencing price spike events, iii) handle the *multimodal* nature of the LMP's probability distribution, and iv) relax the LICQ regularity condition, an assumption that is usually required in the relevant literature [209, 19, 111, 20]. Our results are derived in the case of Gaussian fluctuations, and are validated numerically on the IEEE 14-bus test case.

Chapter outline: The chapter is structured as follows. A rigorous formulation of the problem under consideration is provided in Section 7.1, while a connection to the field of multiparametric programming is established in Section 7.2. Next, in Section 7.3, we derive our main large deviations result linking the event of a rare price spike to the solution of a deterministic optimization problem, which is analyzed in Section 7.4. We illustrate the potential of the proposed methodology in Section 7.5 with a case study on the IEEE 14-bus test case.

7.1 System model and problem formulation

The power grid model shares many similarities to the one described in Chapter 6, and for this reason full details are not repeated here. In particular, we refer to Section 6.1.1 for the definition of the sets $\mathcal{N}_g, \mathcal{N}_w, \mathcal{N}_d$, where $|\mathcal{N}_g| = n_g$, $|\mathcal{N}_w| = n_w$, $|\mathcal{N}_d| = n_d$ and $n_g + n_w = n$ is the total number of nodes in the grid.

We denote the vectors of conventional generation, renewable generation, and demand, as the vectors $\mathbf{g} \in \mathbb{R}_+^{\mathcal{N}_g}$, $\mathbf{w} \in \mathbb{R}_+^{\mathcal{N}_w}$, and $\mathbf{d} \in \mathbb{R}_+^{\mathcal{N}_d}$, respectively. [1] To simplify notation, we extend the vectors $\mathbf{g}, \mathbf{w}, \mathbf{d}$ to n-dimensional vectors

[1] The notation $x \in \mathbb{R}^A$ indicates that the entries in the $|A|$-dimensional vector x are indexed by the set A.

$\tilde{\mathbf{g}}, \tilde{\mathbf{w}}, \tilde{\mathbf{d}} \in \mathbb{R}^n$ by setting $\tilde{g}_i = 0$ whenever $i \notin \mathcal{N}_g$ and $\tilde{g}_i = g_i$ otherwise, and similarly for $\tilde{\mathbf{w}}$ and $\tilde{\mathbf{d}}$. The vectors of *net power injections* and *power flows* are denoted by $\mathbf{p} := \tilde{\mathbf{g}} + \tilde{\mathbf{w}} - \tilde{\mathbf{d}} \in \mathbb{R}^n$ and $\mathbf{f} \in \mathbb{R}^m$, respectively.

As in Chapter 6, the vector of nodal prices $\mathbf{LMP} \in \mathbb{R}^n$ (see Section 1.3.2) is calculated by solving a DC-OPF (see Section 1.3.1), which can be formulated as the following quadratic optimization problem:

$$\min_{\mathbf{g} \in \mathbb{R}^{n_g}} \sum_{i=1}^{n_g} J_i(g_i) = \frac{1}{2}\mathbf{g}^\top \mathbf{H}\mathbf{g} + \mathbf{h}^\top \mathbf{g} \tag{7.2}$$

$$\text{s.t.} \quad \mathbf{1}^\top(\tilde{\mathbf{g}} + \tilde{\mathbf{w}} - \tilde{\mathbf{d}}) = 0 \qquad : \lambda_{\text{en}} \tag{7.3}$$

$$\underline{\mathbf{f}} \leq \widehat{\mathbf{V}}(\tilde{\mathbf{g}} + \tilde{\mathbf{w}} - \tilde{\mathbf{d}}) \leq \bar{\mathbf{f}} \qquad : \boldsymbol{\mu}^-, \boldsymbol{\mu}^+ \tag{7.4}$$

$$\underline{\mathbf{g}} \leq \tilde{\mathbf{g}} \leq \bar{\mathbf{g}} \qquad : \boldsymbol{\tau}^-, \boldsymbol{\tau}^+ \tag{7.5}$$

where the variables are defined as follows:

$\mathbf{H} \in \mathbb{R}^{n_g \times n_g}$	diagonal positive definite matrix, defining the quadratic part of the objective function;
$\mathbf{h} \in \mathbb{R}^{n_g}$	n_g-dimensional vector, defining the linear part of the objective function;
$\widehat{\mathbf{V}} \in \mathbb{R}^{m \times n}$	PTDF matrix (see Section 1.2.2, Eq. (1.20));
$\underline{\mathbf{f}}, \bar{\mathbf{f}} \in \mathbb{R}^m$	vector of lower/upper transmission line limits;
$\underline{\mathbf{g}}, \bar{\mathbf{g}} \in \mathbb{R}^{n_g}$	vector of lower/upper generation constraints;
$\lambda_{\text{en}} \in \mathbb{R}$	dual variable of the energy balance constraint;
$\boldsymbol{\mu}^-, \boldsymbol{\mu}^+ \in \mathbb{R}^m_+$	dual variables of the transmission line constraints (1.26),
$\boldsymbol{\tau}^-, \boldsymbol{\tau}^+ \in \mathbb{R}^{n_g}_+$	dual variables of the generation constraints (1.27);
$\mathbf{1} \in \mathbb{R}^n$	is a n-dimensional vector of ones.

We also denote by $\mathbf{J}(\mathbf{g}) := \sum_{i=1}^n J_i(g_i)$ the aggregated generation cost. The OPF problem in Eqs. (7.2) - (7.5), besides providing the unique[2] *optimal dispatch* $\mathbf{g}^*$, gives as a byproduct the vector of nodal prices $\mathbf{LMP} \in \mathbb{R}^n$, which is related to the dual solution of the DC-OPF as described in Section 1.3.2. In terms of the parameters in Eqs. (7.2) - (7.5), the LMP vector can be expressed as (see Section 6.1.2)

$$\mathbf{LMP} = \lambda_{\text{en}}\mathbf{1} + \widehat{\mathbf{V}}^\top \boldsymbol{\mu} \in \mathbb{R}^n, \tag{7.6}$$

where $\boldsymbol{\mu} = \boldsymbol{\mu}^- - \boldsymbol{\mu}^+$.

[2]The optimal generation dispatch $\mathbf{g}^*$, when it exists, is unique because $\mathbf{H}$ is definite positive, and thus the OPF is a strictly convex optimization problem.

7.1.1 Problem statement

In the same spirit as of Chapter 6, we adopt a *functional* perspective, i.e., we view the uncontrollable generator as a random parameter, or input, of the OPF. In particular, we are interested in a setting where the objective function, PTDF matrix, nodal demand $\mathbf{d}$, line limits and generation constraints are assumed to be known and fixed. Conversely, the uncontrollable generation $\mathbf{w}$ corresponds to a *variable parameter*

$$\boldsymbol{\theta} = \mathbf{w} \in \mathbb{R}^{n_w}, \tag{7.7}$$

of the problem, upon which the solution of the OPF problem in Eqs. (7.2) - (7.5) (henceforth referred to as OPF($\boldsymbol{\theta}$)), and thus the LMP vector, depend.[3]

In other words, the LMP vector is a deterministic function of $\boldsymbol{\theta}$

$$\mathbb{R}^{n_w} \supseteq \boldsymbol{\Theta} \ni \boldsymbol{\theta} \to \mathbf{LMP}(\boldsymbol{\theta}) \in \mathbb{R}^n, \tag{7.8}$$

where $\boldsymbol{\Theta} \subseteq \mathbb{R}^{n_w}$ is the *feasible parameter space* of the OPF, i.e., the set of parameters $\boldsymbol{\theta}$ such that OPF($\boldsymbol{\theta}$) is feasible. In particular, we model $\boldsymbol{\theta}$ as a non-degenerate multivariate Gaussian random vector [4]

$$\boldsymbol{\theta}_\varepsilon \sim \mathcal{N}_{n_w}(\boldsymbol{\mu}_{\boldsymbol{\theta}}, \varepsilon \boldsymbol{\Sigma}_{\boldsymbol{\theta}}), \tag{7.9}$$

where $\varepsilon > 0$ quantifies the magnitude of the noise. The mean $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ of the random vector $\boldsymbol{\theta}$ is interpreted as the expected, or nominal, realization of renewable generation for the considered time interval, and we assume that it lies inside the interior of the feasible parameter set, i.e. $\boldsymbol{\mu}_{\boldsymbol{\theta}} \in \mathring{\boldsymbol{\Theta}}$. Furthermore, we assume that $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ is a known positive definite matrix, and consider the regime where $\varepsilon \to 0$. In view of the mapping (7.8), $\mathbf{LMP}$ is a n-dimensional random vector whose distribution depends on that of $\boldsymbol{\theta}$, and on the deterministic mapping $\boldsymbol{\theta} \to \mathbf{LMP}(\boldsymbol{\theta})$. We assume that the $\mathbf{LMP}$ vector corresponding to the expected renewable generation $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ is such that

$$\mathbf{LMP}(\boldsymbol{\mu}_{\boldsymbol{\theta}}) \in \mathring{\Pi}, \tag{7.10}$$

where $\Pi := \prod_{i=1}^{n} [\alpha_i^-, \alpha_i^+]$, $\boldsymbol{\alpha}^-, \boldsymbol{\alpha}^+ \in \mathbb{R}^n$ are vectors of price thresholds. Here and in the following, the notation $\mathring{A}$ denotes the interior of the set A.

[3] We introduce the notation with $\boldsymbol{\theta}$ in order to describe a more general framework. For example, the parameter $\boldsymbol{\theta}$ may model additional quantities, such as variable demand, as in Chapter 6.

[4] Technically, $\boldsymbol{\theta}_\varepsilon$ is truncated by the feasibility set $\boldsymbol{\Theta}$. From the viewpoint of the large deviations approach, however, non-feasible realizations of $\boldsymbol{\theta}_\varepsilon$ do not constitute a problem, since the large deviations result in Eq. (7.19) involves solving an optimization problem defined on the set $Y \subseteq \boldsymbol{\Theta}$.

We are interested in the event of *anomalous price fluctuations* (or *spikes*), defined as

$$Y = Y(\boldsymbol{\alpha}^-, \boldsymbol{\alpha}^+) = \left\{\boldsymbol{\theta} \in \boldsymbol{\Theta} \,:\, \mathbf{LMP}(\boldsymbol{\theta}) \notin \prod_{i=1}^{n} [\alpha_i^-, \alpha_i^+]\right\} \tag{7.11}$$

$$= \bigcup_{i=1}^{n} \{\boldsymbol{\theta} \in \boldsymbol{\Theta} \,:\, \mathrm{LMP}_i(\boldsymbol{\theta}) < \alpha_i^- \text{ or } \mathrm{LMP}_i(\boldsymbol{\theta}) > \alpha_i^+\}, \tag{7.12}$$

which, in view of Eq. (7.10) and the regime $\varepsilon \to 0$, is a *rare event*. Without loss of generality, we only consider thresholds $\boldsymbol{\alpha}^-, \boldsymbol{\alpha}^+$ such that the event $Y(\boldsymbol{\alpha}^-, \boldsymbol{\alpha}^+)$ has a non-empty interior in $\mathbb{R}^{n_w}$. Otherwise, the fact that $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ is non degenerate would imply $\mathbb{P}\big(Y(\boldsymbol{\alpha}^-, \boldsymbol{\alpha}^+)\big) = 0$.

We observe that the above formulation of a spike event is quite general, and can cover different application scenarios, as we now describe. For example, if $\alpha = \alpha_i^+ = -\alpha_i^- > 0$ for all i, then the spike event reads

$$Y(\alpha) = \{\boldsymbol{\theta} \in \Theta \,:\, \|\mathbf{LMP}\|_\infty = \max_{i=1,\dots,n} |\mathrm{LMP}_i| > \alpha\},$$

modeling the occurrence of a price spike with magnitude greater than a prescribed value. On the other hand, if we define $\boldsymbol{\alpha}^- = \mathbf{LMP}(\boldsymbol{\mu}_{\boldsymbol{\theta}}) - \boldsymbol{\beta}$ and $\boldsymbol{\alpha}^+ = \mathbf{LMP}(\boldsymbol{\mu}_{\boldsymbol{\theta}}) + \boldsymbol{\beta}$, for $\boldsymbol{\beta} \in \mathbb{R}^n_+$, the spike event

$$Y(\boldsymbol{\beta}) = \bigcup_{i=1}^{n} \{\boldsymbol{\theta} \in \boldsymbol{\Theta} \,:\, |\mathrm{LMP}_i - \mathrm{LMP}_i(\boldsymbol{\mu}_{\boldsymbol{\theta}})| > \beta_i\},$$

models the event of any LMP_i deviating from its nominal value $\mathrm{LMP}_i(\boldsymbol{\mu}_{\boldsymbol{\theta}})$ more than $\beta_i > 0$. Moreover, by setting $\boldsymbol{\alpha}^- = \mathbf{LMP}(\boldsymbol{\mu}_{\boldsymbol{\theta}}) - \boldsymbol{\beta}^-$ and $\boldsymbol{\alpha}^+ = \mathbf{LMP}(\boldsymbol{\mu}_{\boldsymbol{\theta}}) + \boldsymbol{\beta}^+$, $\boldsymbol{\beta}^-, \boldsymbol{\beta}^+ \in \mathbb{R}^{m,+}$ and $\boldsymbol{\beta}^- \neq \boldsymbol{\beta}^+$, we can weigh differently negative and positive deviations from the nominal values.

We remark that *negative* price spikes are also of interest [71, 75] and can be covered in our framework, by choosing the threshold vectors $\boldsymbol{\alpha}^-, \boldsymbol{\alpha}^+$ accordingly. Finally, note that we can study price spikes at a more granular level by restricting the union in Eq. (7.12) to a particular subset of nodes $\widetilde{\mathcal{N}} \subseteq \mathcal{N}$.

7.2 Multiparametric programming

As already observed multiple times throughout this thesis (see Section 1.1 and Section 6.1.3, for instance), the LMPs can be thought as deterministic function of the parameter $\boldsymbol{\theta}$. Therefore, in order to study the distribution of

the random vector $\mathbf{LMP}$, we need to investigate the structure of the mapping $\boldsymbol{\theta} \to \mathbf{LMP}(\boldsymbol{\theta})$.

In the same way as in Chapter 6, we first rewrite the problem OPF($\boldsymbol{\theta}$) in Eqs. (7.2) - (7.5) as a standard Multiparametric Quadratic Program (MPQ), yielding the equivalent formulation

$$\min_{\mathbf{g}\in\mathbb{R}^{n_g}} \quad \frac{1}{2}\mathbf{g}^\top \mathbf{H}^\top \mathbf{g} + \mathbf{g}^\top \mathbf{h} \tag{7.13}$$

$$\text{s.t.} \quad \mathbf{A}\mathbf{g} \leq \mathbf{b} + \mathbf{E}\boldsymbol{\theta}, \tag{7.14}$$

where $\mathbf{A} \in \mathbb{R}^{(2+2m+2n_g)\times n_g}, \mathbf{E} \in \mathbb{R}^{(2+2m+2n_g)\times n_w}, \mathbf{b} \in \mathbb{R}^{(2+2m+2n_g)}$ are suitably defined matrices and vectors, given below:

$$\mathbf{A} = \begin{bmatrix} \mathbf{1}_{n_g}^\top \\ -\mathbf{1}_{n_g}^\top \\ \widehat{\mathbf{V}}_{\mathcal{N}_g} \\ -\widehat{\mathbf{V}}_{\mathcal{N}_g} \\ \mathbf{I}_{n_g} \\ -\mathbf{I}_{n_g} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} \mathbf{1}^\top \mathbf{d} \\ -\mathbf{1}^\top \mathbf{d} \\ \widehat{\mathbf{V}}_{\mathcal{N}_d}\mathbf{d} + \bar{\mathbf{f}} \\ -\widehat{\mathbf{V}}_{\mathcal{N}_d}\mathbf{d} - \underline{\mathbf{f}} \\ \bar{\mathbf{g}} \\ -\underline{\mathbf{g}} \end{bmatrix}, \quad \mathbf{E} = \begin{bmatrix} -\mathbf{1}_{n_w}^\top \\ \mathbf{1}_{n_w}^\top \\ -\widehat{\mathbf{V}}_{\mathcal{N}_w} \\ \widehat{\mathbf{V}}_{\mathcal{N}_w} \\ \mathbf{0}_{n_w} \\ \mathbf{0}_{n_w} \end{bmatrix}. \tag{7.15}$$

Here, for $k \in \mathbb{N}$, we denote by $\mathbf{1}_k, \mathbf{0}_k \in \mathbb{R}^k$ and $\mathbf{I}_k \in \mathbb{R}^{k\times k}$ the vector of ones, zeros, and the identity matrix of dimension k, respectively. Moreover, $\widehat{\mathbf{V}}_{\mathcal{N}_g} \in \mathbb{R}^{m\times n_g}$ and $\widehat{\mathbf{V}}_{\mathcal{N}_w} \in \mathbb{R}^{m\times n_w}$ denote the submatrices of $\widehat{\mathbf{V}}$ obtained by selecting only the columns corresponding to nodes in $\mathcal{N}_g$ and $\mathcal{N}_w$, respectively.

In the remainder of this chapter, we make use of the concepts of optimal partition and LICQ condition, which have already been introduced in Definitions 6.1 and 6.2. We recall that, for a given parameter $\boldsymbol{\theta}$, the associated optimal partition $\mathcal{B}(\boldsymbol{\theta})$ consists of the indices of binding constraints of OPF($\boldsymbol{\theta}$) in Eqs. (7.13) - (7.14).

Since there is always at least one binding constraint, namely $i = 1$, corresponding to the power balance constraint (see also Remark 6.1 in Chapter 6), we can write $|\mathcal{B}| = 1 + |\mathcal{B}'|$, where $\mathcal{B}' \subseteq \{3, \dots, 2 + 2n_g + 2m\}$ contains the indexes of binding constraints corresponding to line and generator limits. Let $\tilde{n}_g, \tilde{m}$ denote, respectively, the number of binding generator and line limit constraints, so that $\mathcal{B}' = \tilde{n}_g + \tilde{m}$. Since line and generation limits cannot be binding both on the positive and negative sides, we have that $|\mathcal{B}'| \leq n_g + m$.

Moreover, it is observed in [209] that the row rank of $\mathbf{A}_\mathcal{B}$ is equal to $\min(1 + \tilde{n_g} + \tilde{m}, n_g)$. Therefore, the LICQ condition is equivalent to

$$1 + \tilde{n_g} + \tilde{m} \leq n_g. \tag{7.16}$$

A standard result in Multiparametric Programming Theory (presented earlier as Theorem 6.1 in Section 6.2, and proved in [181]) states that the feasible parameter space $\boldsymbol{\Theta}$ can be covered by the union of a finite number of convex polytopes $\boldsymbol{\Theta}_k$ (called *critical regions*) $k = 1, \dots, M$, and that there exist M affine maps defined in the interiors of the critical regions

$$\mathring{\boldsymbol{\Theta}}_k \ni \boldsymbol{\theta} \to \mathbf{LMP}\big|_{\mathring{\boldsymbol{\Theta}}_k}(\boldsymbol{\theta}) = \tilde{\mathbf{C}}^{(k)}\boldsymbol{\theta} + \tilde{\mathbf{c}}^{(k)},\ k = 1 \dots, M,$$

where $\tilde{\mathbf{C}}^{(k)}, \tilde{\mathbf{c}}^{(k)}$ are suitably defined matrices and vectors. Moreover, if LICQ holds for every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, then the maps agree on the intersections between the regions $\boldsymbol{\Theta}_k$'s, resulting in an overall continuous map

$$\boldsymbol{\Theta} \ni \boldsymbol{\theta} \to \mathbf{LMP}(\boldsymbol{\theta}) \in \mathbb{R}^n. \tag{7.17}$$

7.2.1 Relaxing the LICQ assumption

One of the assumptions of Theorem 6.1, which is standard in the field [209, 19, 111, 20], is that the LICQ condition holds for every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. In particular, this means that LICQ holds in the interior of two neighboring regions, which we denote as $\mathring{\boldsymbol{\Theta}}_i$ and $\mathring{\boldsymbol{\Theta}}_j$. Let $\mathcal{H}$ be the hyperplance separating $\mathring{\boldsymbol{\Theta}}_i$ and $\mathring{\boldsymbol{\Theta}}_j$. The fact that LICQ holds at $\mathring{\boldsymbol{\Theta}}_i$ implies that, if $\{i_1, \dots, i_q\}$ are the binding constraints at optimality in the OPF for $\boldsymbol{\theta} \in \mathring{\boldsymbol{\Theta}}_i$, then in view of Eq. (7.16) we have $q \leq n_g$, where n_g is the number of decision variables in the OPF (i.e., the number of generators).

Requiring LICQ to hold everywhere means that, in particular, it must hold in the common facet between regions. As we move from $\mathring{\boldsymbol{\Theta}}_k$ on to the common facet $\mathcal{F} = \boldsymbol{\Theta}_i \cap \mathcal{H}$ between regions $\boldsymbol{\Theta}_i$ and $\boldsymbol{\Theta}_j$, which has dimension $n_w - 1$, there could be an additional constraint becoming active (coming from the neighboring region $\boldsymbol{\Theta}_j$), and therefore the LICQ condition implies $q + 1 \leq n_g$.

In general, critical regions can intersect in faces of dimensions $1, \dots, n_w - 1$, and enforcing LICQ to hold on all these faces could imply the overly-conservative assumption $q + n_w - 1 \leq n_g$.

In what follows, we relax the assumptions of Theorem 6.1 by allowing LICQ to be violated on the union of these lower-dimensional faces

$$\boldsymbol{\Theta}_\circ := \boldsymbol{\Theta} \setminus \bigcup_{k=1}^{M} \mathring{\boldsymbol{\Theta}}_k.$$

Since this union has zero n_w-dimensional Lebesgue measure, the event $\boldsymbol{\theta} \in \boldsymbol{\Theta}_\circ$ rarely happens in practice, and thus is usually ignored in the literature, but it does cause technical issues that we now address.

If LICQ is violated on $\boldsymbol{\theta} \in \boldsymbol{\Theta}_\circ$, the Lagrange multipliers of the OPF, and thus the LMP, need not be unique. Therefore, the map $\boldsymbol{\theta} \to \mathbf{LMP}(\boldsymbol{\theta})$ is not properly defined on $\boldsymbol{\Theta}_\circ$. In order to extend the map from $\bigcup_{k=1}^{M} \mathring{\boldsymbol{\Theta}}_k$ to the full feasible parameter space $\boldsymbol{\Theta}$, we incorporate a tie-breaking rule to consistently choose between the possible LMPs.

Following [180], we break ties by using the lexicographic order. That is, if $\{\mathbf{LMP}^1(\boldsymbol{\theta}), \dots, \mathbf{LMP}^d(\boldsymbol{\theta})\}$ denotes the set of LMPs corresponding to a parameter $\boldsymbol{\theta} \in \mathcal{F}$, we first choose those with maximum first coordinate $\{\mathbf{LMP}^j(\boldsymbol{\theta})\}_{j \in \mathcal{K}_1}$, with $\mathcal{K}_1 = \arg\max_{h=1,\dots,d} \mathbf{LMP}_1^h(\boldsymbol{\theta})$. Then, from this subset, we choose the LMP vectors with maximum second coordinate, and so forth.

This choice defines the LMP function over the whole feasible parameter space $\boldsymbol{\Theta}$, but may introduce jump discontinuities on the zero-measure set $\boldsymbol{\Theta}_\circ$. In the next section, we address this technicality and formally derive our main large deviations result.

7.3 Large deviations results

Proposition 7.1. *Let $\boldsymbol{\theta}_\varepsilon \sim \mathcal{N}_{n_\theta}(\boldsymbol{\mu}_{\boldsymbol{\theta}}, \varepsilon \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$ be a family of nondegenerate n_θ-dimensional Gaussian r.v.'s indexed by $\varepsilon > 0$. Assume that the LICQ condition is satisfied for all $\boldsymbol{\theta} \in \boldsymbol{\Theta} \setminus \boldsymbol{\Theta}_\circ$. Consider the event*

$$Y = Y(\boldsymbol{\alpha}^-, \boldsymbol{\alpha}^+) = \bigcup_{i=1}^{n} \{\boldsymbol{\theta} \in \boldsymbol{\Theta} \,:\, \mathrm{LMP}_i(\boldsymbol{\theta}) \notin [\alpha_i^-, \alpha_i^+]\},$$

defined in Eq. (7.11)*, assume that the interior of Y is not empty,* [5] *and that*

$$\mathbf{LMP}(\boldsymbol{\mu}_{\boldsymbol{\theta}}) \in \mathring{\Pi}, \quad \Pi := \prod_{i=1}^{n} [\alpha_i^-, \alpha_i^+]. \tag{7.18}$$

Then, the family of random vectors $\{\boldsymbol{\theta}_\varepsilon\}_{\varepsilon > 0}$ satisfies

$$\lim_{\varepsilon \to 0} \varepsilon \log \mathbb{P}(\boldsymbol{\theta}_\varepsilon \in Y) = - \inf_{\boldsymbol{\theta} \in Y} I(\boldsymbol{\theta}), \tag{7.19}$$

where $\mathcal{I}(\boldsymbol{\theta}) = \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}})^\top \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}})$..

Proof. For notational compactness, in the rest of the proof we will write Y without making explicit its dependence on $(\boldsymbol{\alpha}^-, \boldsymbol{\alpha}^+)$. Defining $Z := \bigcup_{i=1}^{n} \{\boldsymbol{\theta} \in$

[5] If $\mathring{Y} = \emptyset$, then trivially $\mathbb{P}(\boldsymbol{\theta}_\varepsilon \in Y) = 0$.

$\mathbb{R}^{n_\theta} : \mathrm{LMP}_i(\boldsymbol{\theta}) \notin [\alpha_i^-, \alpha_i^+]\}$, the event Y can be decomposed as the disjoint union $Y = Y_* \cup Y_\circ$, where

$$Y_* = \bigcup_{k=1}^{M} \mathring{\boldsymbol{\Theta}}_k \cap Z, \quad Y_\circ = \boldsymbol{\Theta}_\circ \cap Z, \tag{7.20}$$

and $Y_\circ \subseteq \boldsymbol{\Theta}_\circ = \boldsymbol{\Theta} \setminus \bigcup_{k=1}^{M} \mathring{\boldsymbol{\Theta}}_k$ is a zero-measure set.

As $\boldsymbol{\theta}_\varepsilon$ is nondegenerate, it has a density f with respect to the n_w-dimensional Lebesgue measure in $\mathbb{R}_w^n$. Since the n_w-dimensional Lebesgue measure of $Y_\circ$ is zero, we have

$$\mathbb{P}(\boldsymbol{\theta}_\varepsilon \in Y_\circ) = \int_{\mathbf{x} \in Y_\circ} f(\mathbf{x}) d\mathbf{x} = 0$$

and $\mathbb{P}(\boldsymbol{\theta}_\varepsilon \in Y) = \mathbb{P}(\boldsymbol{\theta}_\varepsilon \in Y_*)$. As a consequence, we can restrict our analysis to the event Y_*. Thanks to Cramer's theorem in $\mathbb{R}^{n_w}$ [55], we have

$$\begin{aligned} -\inf_{\boldsymbol{\theta} \in \mathring{Y}_*} I(\boldsymbol{\theta}) &\leq \liminf_{\varepsilon \to 0} \varepsilon \log\left(\mathbb{P}(\boldsymbol{\theta}_\varepsilon \in Y_*)\right) && (7.21) \\ &\leq \limsup_{\varepsilon \to 0} \varepsilon \log\left(\mathbb{P}(\boldsymbol{\theta}_\varepsilon \in Y_*)\right) \leq -\inf_{\boldsymbol{\theta} \in \overline{Y_*}} I(\boldsymbol{\theta}), && (7.22) \end{aligned}$$

where $I(\boldsymbol{\theta})$ is the Legendre transform of the log-moment generating function of $\boldsymbol{\theta}_\varepsilon$. It is well-known (see, for example, [182]) that when $\boldsymbol{\theta}_\varepsilon$ is Gaussian then $\mathcal{I}(\boldsymbol{\theta}) = (\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}})^\top \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}})$. In order to prove (7.19), it remains to be shown that

$$\inf_{\boldsymbol{\theta} \in \mathring{Y}_*} I(\boldsymbol{\theta}) = \inf_{\boldsymbol{\theta} \in \overline{Y_*}} I(\boldsymbol{\theta}). \tag{7.23}$$

Thanks to the continuity of the maps $\mathbf{LMP}|_{\mathring{\boldsymbol{\Theta}}_k}$, the set Y_* is open, since

$$\begin{aligned} Y_* &= \bigcup_{k=1}^{M} \mathring{\boldsymbol{\Theta}}_k \cap Z && (7.24) \\ &= \bigcup_{k=1}^{M} \Big(\mathring{\boldsymbol{\Theta}}_k \cap \bigcup_{i=1}^{n} \{ \mathrm{LMP}_i|_{\mathring{\boldsymbol{\Theta}}_k}(\boldsymbol{\theta}) \notin [\alpha_i^-, \alpha_i^+]\}\Big) && (7.25) \\ &= \bigcup_{k=1}^{M} \Big(\mathring{\boldsymbol{\Theta}}_k \cap \bigcup_{i=1}^{n} \{\tilde{\mathbf{C}}_i^{(k)} \boldsymbol{\theta} + \tilde{\mathbf{c}}_i^{(k)} \notin [\alpha_i^-, \alpha_i^+]\}\Big) && (7.26) \\ &= \bigcup_{k=1}^{M} \bigcup_{i=1}^{n} \Big(\mathring{\boldsymbol{\Theta}}_k \cap (\{\tilde{\mathbf{C}}_i^{(k)} \boldsymbol{\theta} + \tilde{\mathbf{c}}_i^{(k)} < \alpha_i^-\} \cup \{\tilde{\mathbf{C}}^{(k)} \boldsymbol{\theta} + \tilde{\mathbf{c}}^{(k)} > \alpha_i^+\})\Big). && (7.27) \end{aligned}$$

Therefore, $\overset{\circ}{Y}_* = Y_*$ and $\overline{\overset{\circ}{Y}_*} = \overline{Y_*} \supseteq Y_*$. Since the rate function $I(\boldsymbol{\theta})$ is continuous, Eq. (7.23) follows. □

Proposition 7.1 allows us approximate the probability of a price spike, for small ε, as

$$\mathbb{P}(\boldsymbol{\theta}_\varepsilon \in Y) \approx \exp\Big(\frac{-\inf_{\boldsymbol{\theta}\in Y} I(\boldsymbol{\theta})}{\varepsilon}\Big), \tag{7.28}$$

as it was done in Chapters 2 and 4 in the context of studying the event of transmission line failures. The minimizer of the optimization problem (7.19) corresponds to the most likely realization of uncontrollable generation that leads to the rare event.

Furthermore, the structure of the problem (7.19) allows us to efficiently rank nodes in terms of their likelihood to experience a price spike, as we illustrate in Section 7.5.

7.4 Solving the optimization problem

In view of Proposition 7.1, in order to study $\lim_{\varepsilon\to 0} \varepsilon \log \mathbb{P}(\boldsymbol{\theta}_\varepsilon \in Y(\boldsymbol{\alpha}^-, \boldsymbol{\alpha}^+))$ we need to solve the deterministic optimization problem $\inf_{\boldsymbol{\theta}\in Y_*} I(\boldsymbol{\theta})$. The latter, in view of Theorem 6.1 and the definition of Y_*, is equivalent to

$$\begin{aligned}\inf_{\boldsymbol{\theta}\in Y_*} I(\boldsymbol{\theta}) &= \min_{k=1,\dots,M} \inf_{\boldsymbol{\theta}\in \mathring{\Theta}_k \cap Z} I(\boldsymbol{\theta}) \\ &= \min_{i=1,\dots,n} \min_{k=1,\dots,M} \inf_{\boldsymbol{\theta}\in\mathring{\Theta}_k, \tilde{\mathbf{C}}_i^{(k)}\boldsymbol{\theta}+\tilde{\mathbf{c}}_i^{(k)} \notin [\alpha_i^-, \alpha_i^+]} I(\boldsymbol{\theta}).\end{aligned}$$

This amounts to solving at most nM quadratic optimization problems of the form

$$\inf_{\boldsymbol{\theta}\in T_{i,k}} I(\boldsymbol{\theta})$$

for $i = 1, \dots, n$, $k = 1, \dots, M$, where, for $i = 1, \dots, n$, $k = 1, \dots, M$,

$$T_{i,k} = T_{i,k}^- \sqcup T_{i,k}^+,$$

$$T_{i,k}^- = \mathring{\Theta}_k \cap \{\tilde{\mathbf{C}}_i^{(k)}\boldsymbol{\theta} + \tilde{\mathbf{c}}_i^{(k)} < \alpha_i^-\}, \quad T_{i,k}^+ = \mathring{\Theta}_k \cap \{\tilde{\mathbf{C}}_i^{(k)}\boldsymbol{\theta} + \tilde{\mathbf{c}}_i^{(k)} > \alpha_i^+\}.$$

In the rest of this section, we show how we can significantly reduce the number of optimization problems that need to be solved by exploiting the geometric structure of the problem. First, since

$$\inf_{\boldsymbol{\theta}\in Y_*} I(\boldsymbol{\theta}) = \min_{i=1,\dots,n} \inf_{\boldsymbol{\theta}\in\bigcup_{k=1}^M T_{i,k}} I(\boldsymbol{\theta}),$$

we fix $i = 1 \dots, n$ and consider the sub-problems

$$\inf_{\boldsymbol{\theta} \in \bigcup_{k=1}^{M} T_{i,k}} I(\boldsymbol{\theta}) = \min \Big\{ \inf_{\boldsymbol{\theta} \in \bigcup_{k=1}^{M} T_{i,k}^{-}} I(\boldsymbol{\theta}), \inf_{\boldsymbol{\theta} \in \bigcup_{k=1}^{M} T_{i,k}^{+}} I(\boldsymbol{\theta}) \Big\}. \tag{7.29}$$

The reason why we want to solve the problems in Eq. (7.29) individually for every i is because we are not only interested in studying the overall event Y, but also in the more granular events of node-specific price spikes. For example, this would allow us to rank the nodes in terms of their likelihood of experiencing a price spike (see Section 7.5). Define

$$\begin{aligned}
L_{(i,k)}^{-} &:= \boldsymbol{\Theta}_k \cap (T_{i,k}^{-})^{\complement} = \boldsymbol{\Theta}_k \cap \{\tilde{\mathbf{C}}_i^{(k)} \boldsymbol{\theta} + \tilde{\mathbf{c}}_i^{(k)} \geq \alpha_i^{-}\}, \\
L_{(i,k)}^{+} &:= \boldsymbol{\Theta}_k \cap (T_{i,k}^{+})^{\complement} = \boldsymbol{\Theta}_k \cap \{\tilde{\mathbf{C}}_i^{(k)} \boldsymbol{\theta} + \tilde{\mathbf{c}}_i^{(k)} \leq \alpha_i^{+}\}, \\
L_i^{-} &:= \bigcup_{k=1}^{M} L_{i,k}^{-} = \boldsymbol{\Theta} \cap \{\mathrm{LMP}_i \geq \alpha_i^{-}\}, \\
L_i^{+} &:= \bigcup_{k=1}^{M} L_{i,k}^{+} = \boldsymbol{\Theta} \cap \{\mathrm{LMP}_i \leq \alpha_i^{+}\}.
\end{aligned}$$

and consider the partition of the sets L_i^{+} and L_i^{-} into disjoint closed connected components, i.e.,

$$L_i^{-} = \bigsqcup_{\ell \in \text{conn. comp. of } L_i^{-}} W_{\ell}^{(i,-)}, \; L_i^{+} = \bigsqcup_{\ell \in \text{conn. comp. of } L_i^{+}} W_{\ell}^{(i,+)}, \tag{7.30}$$

and let $W_{\ell^{-*}}^{(i,-)}, W_{\ell^{+*}}^{(i,+)}$ be the components containing $\boldsymbol{\mu}_{\boldsymbol{\theta}}$. Since $\partial(A \cup B) = \partial A \cup \partial B$ if $\overline{A} \cap B = A \cap \overline{B} = \emptyset$, the boundary $\partial L_i^{+} = \bigsqcup_{\ell \in \mathcal{F}_i^{+}} \partial W_{\ell}^{(i,+)}$ is the union of the set of parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ such that $\mathbf{LMP}(\boldsymbol{\theta}) = \alpha_i^{+}$ with, possibly, a subset of the boundary of $\boldsymbol{\Theta}$ (and similarly for ∂L_i^{-}).

As stated by Proposition 7.2, we show that, in order to solve the two problems in the right hand side of Eq. (7.29) we need to look only at the boundaries $\partial W_{\ell^*}^{(i,-)}, \partial W_{\ell^*}^{(i,+)}$.

Proposition 7.2. *Under the same assumptions of Theorem 6.1, we have*

$$\inf_{\boldsymbol{\theta} \in \bigcup_{k=1}^{M} T_{i,k}^{+}} I(\boldsymbol{\theta}) = \inf_{\boldsymbol{\theta} \in \partial \bigcup_{k=1}^{M} T_{i,k}^{+}} I(\boldsymbol{\theta}), \quad \inf_{\boldsymbol{\theta} \in \bigcup_{k=1}^{M} T_{i,k}^{-}} I(\boldsymbol{\theta}) = \inf_{\boldsymbol{\theta} \in \partial \bigcup_{k=1}^{M} T_{i,k}^{-}} I(\boldsymbol{\theta}). \tag{7.31}$$

Moreover,

$$\inf_{\boldsymbol{\theta}\in\bigcup_{k=1}^{M} T_{i,k}^{+}} I(\boldsymbol{\theta}) = \inf_{\boldsymbol{\theta}\in\partial W_{\ell^*+}^{(i,+)}} I(\boldsymbol{\theta}), \qquad \inf_{\boldsymbol{\theta}\in\bigcup_{k=1}^{M} T_{i,k}^{-}} I(\boldsymbol{\theta}) = \inf_{\boldsymbol{\theta}\in\partial W_{\ell^*-}^{(i,-)}} I(\boldsymbol{\theta}). \tag{7.32}$$

Proof. First note that the rate function $I(\boldsymbol{\theta})$ is a (strictly) convex function, since $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ is positive definite. Since $\bigcup_{k=1}^{M} T_{i,k}^{+}$ is open and $I(\boldsymbol{\theta})$ is a continuous function, it holds that

$$\inf_{\boldsymbol{\theta}\in\bigcup_{k=1}^{M} T_{i,k}^{+}} I(\boldsymbol{\theta}) = \inf_{\boldsymbol{\theta}\in\overline{\bigcup_{k=1}^{M} T_{i,k}^{+}}} I(\boldsymbol{\theta}).$$

Moreover, since $I(\boldsymbol{\theta})$ is continuous and $\overline{\bigcup_{k=1}^{M} T_{i,k}^{+}}$ compact, the infimum is attained. The fact that $\overline{\bigcup_{k=1}^{M} T_{i,k}^{+}} \supseteq \partial \bigcup_{k=1}^{M} T_{i,k}^{+}$ immediately implies that

$$\inf_{\boldsymbol{\theta}\in\bigcup_{k=1}^{M} T_{i,k}^{+}} I(\boldsymbol{\theta}) \leq \inf_{\boldsymbol{\theta}\in\partial\bigcup_{k=1}^{M} T_{i,k}^{+}} I(\boldsymbol{\theta}).$$

On the other hand, assume by contradiction that

$$\inf_{\boldsymbol{\theta}\in\bigcup_{k=1}^{M} T_{i,k}^{+}} I(\boldsymbol{\theta}) < \inf_{\boldsymbol{\theta}\in\partial\bigcup_{k=1}^{M} T_{i,k}^{+}} I(\boldsymbol{\theta}).$$

In particular, there exists a point $\boldsymbol{\theta}_0$ in the interior of $\bigcup_{k=1}^{M} T_{i,k}^{+}$ such that $I(\boldsymbol{\theta}_0) < I(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \bigcup_{k=1}^{M} T_{i,k}^{+}$. Define, for $t \in [0,1]$, the line segment joining $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$, i.e. $\boldsymbol{\theta}_t = (1-t)\boldsymbol{\mu}_{\boldsymbol{\theta}} + t\boldsymbol{\theta}_0$. Since $\boldsymbol{\theta}_0$ lies in the interior of $\bigcup_{k=1}^{M} T_{i,k}^{+}$, and $\boldsymbol{\mu}_{\boldsymbol{\theta}} \notin \bigcup_{k=1}^{M} T_{i,k}^{+}$, there exist a $0 < t_* < 1$ such that $\boldsymbol{\theta}_t \in \bigcup_{k=1}^{M} T_{i,k}^{+}$ for all $t \in [t_*, 1]$. Due to the convexity of $I(\boldsymbol{\theta})$, and the fact that $I(\boldsymbol{\mu}_{\boldsymbol{\theta}}) = 0$, we have

$$I(\boldsymbol{\theta}_{t_*}) < (1-t_*)I(\boldsymbol{\mu}_{\boldsymbol{\theta}}) + t_* I(\boldsymbol{\theta}_0) = t_* I(\boldsymbol{\theta}_0) < I(\boldsymbol{\theta}_0),$$

thus reaching a contradiction. Hence,

$$\inf_{\boldsymbol{\theta}\in\bigcup_{k=1}^{M} T_{i,k}^{+}} I(\boldsymbol{\theta}) = \inf_{\boldsymbol{\theta}\in\partial\bigcup_{k=1}^{M} T_{i,k}^{+}} I(\boldsymbol{\theta}), \tag{7.33}$$

and the minimum is achieved on $\partial\bigcup_{k=1}^{M} T_{i,k}^{+}$, proving Eq. (7.31).

In view of Eq. (7.31), in order to prove Eq. (7.32) it is enough to show that

$$\inf_{\boldsymbol{\theta}\in\partial\bigcup_{k=1}^{M} T_{i,k}^{+}} I(\boldsymbol{\theta}) = \inf_{\boldsymbol{\theta}\in\partial W_{\ell^*+}^{(i,+)}} I(\boldsymbol{\theta}).$$

Given that the sets $T^+_{i,k} = \mathring{\boldsymbol{\Theta}}_k \cup \{\mathrm{LMP}_i(\boldsymbol{\theta}) > \alpha^+_i\}$, for $k = 1, \dots, M$, are disjoint, the boundary of the union is equal to the union of the boundaries, i.e., $\partial \bigcup_{k=1}^M T^+_{i,k} = \bigcup_{k=1}^M \partial T^+_{i,k}$. Each term $\partial T^+_{i,k}$ is the boundary of the polytope $\overline{T^+}_{i,k} = \boldsymbol{\Theta}_k \cap \{\mathrm{LMP}_i \geq \alpha^+_i\}$, and thus consists of the union of a subset of $\bigcup_{k=1}^M \partial \boldsymbol{\Theta}_k$ (a subset of the union of the facets of the polytope $\boldsymbol{\Theta}_k$) with the segment $\boldsymbol{\Theta}_k \cap \{\mathrm{LMP}_i = \alpha^+_i\}$. As a result, $\partial \bigcup_{k=1}^M T^+_{i,k} I(\boldsymbol{\theta})$ intersects $\partial W^{(i,+)}_{\ell^{*}+}$ in $\boldsymbol{\Theta} \cap \{\mathrm{LMP}_i = \alpha^+_i\}$.

We now show that (i) the minimum is attained at a point $\boldsymbol{\theta}_0$ such that $\mathrm{LMP}_i(\boldsymbol{\theta}_0) = \alpha^+_i$, so that $\boldsymbol{\theta}_0 \in \bigsqcup_{\ell \in \text{conn. comp. of } L^+_i} \partial W^{(i,+)}_\ell$, and (ii) $\boldsymbol{\theta}_0 \in \partial W^{(i,+)}_{\ell^{*}+}$. Assume by contradiction that $\mathrm{LMP}_i(\boldsymbol{\theta}_0) > \alpha^+_i$, and consider the line segment joining $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$, $\boldsymbol{\theta}_t = (1-t)\boldsymbol{\mu}_{\boldsymbol{\theta}} + t\boldsymbol{\theta}_0$, $t \in [0,1]$. The function

$$[0,1] \ni t \to g(t) := \mathrm{LMP}_i(\boldsymbol{\theta}_t) = \mathrm{LMP}_i((1-t)\boldsymbol{\mu}_{\boldsymbol{\theta}} + t\boldsymbol{\theta}_0) \in \mathbb{R},$$

is continuous and such that $g(0) = \mathbf{LMP}(\boldsymbol{\mu}_{\boldsymbol{\theta}}) < \alpha$ and $g(1) = \mathbf{LMP}(\boldsymbol{\theta}_0) > \alpha^*_i$. Thanks to the intermediate value theorem, there exists a $0 < t_* < 1$ such that $g(t_*) = \mathrm{LMP}_i(\boldsymbol{\theta}_{t_*}) = \alpha^*_i$, and

$$I(\boldsymbol{\theta}_{t_*}) < (1-t_*)I(\boldsymbol{\mu}_{\boldsymbol{\theta}}) + t_* I(\boldsymbol{\theta}_0) = t_* I(\boldsymbol{\theta}_0) < I(\boldsymbol{\theta}_0),$$

which is a contradiction, since $\boldsymbol{\theta}_0$ is the minimum. The same argument, based on the convexity of the rate function and the fact that $I(\boldsymbol{\mu}_{\boldsymbol{\theta}}) = 0$, shows that $\boldsymbol{\theta} \in \partial W^{(i,+)}_{\ell^{*},+}$. Lastly, Eq. (7.32) can be derived in the same way. □

Proposition 7.2 shows that in order to solve the problem in Eq. (7.29) we only need to look at the boundaries $\partial W^{(i,+)}_{\ell^*}, \partial W^{(i,-)}_{\ell^*}$. Determining such boundaries is a non-trivial problem, for which dedicated algorithms exist. Such algorithms are beyond the scope of this chapter, and we refer the interested reader to the contour tracing literature and, in particular, to [56].

7.5 Numerics

In this section, we illustrate the potential of our large deviations approach using IEEE 14-bus test case in MATPOWER [210]. This network consists of 14 nodes (each of which houses a load), 6 controllable generators, and 20 lines. As line limits are not included in the test case, we set them as $\bar{\mathbf{f}} = \lambda \bar{\mathbf{f}}^{(\text{planning})}$, where $\bar{\mathbf{f}}^{(\text{planning})} := \gamma_{\text{line}}|\mathbf{f}|$, $\mathbf{f}$ is the solution of a DC-OPF using the data in the test file, and $\gamma_{\text{line}} \geq 1$. We interpret $\bar{\mathbf{f}}^{(\text{planning})}$ as the maximum allowable power flow before the line trips, while the more conservative

$$\bar{\mathbf{f}} = \lambda \bar{\mathbf{f}}^{(\text{planning})}, \tag{7.34}$$

with $1/\gamma_{\text{line}} \leq \lambda \leq 1$, is the operational line limit (see also Section 5.2.1 in Chapter 5). In the rest of this section, we set $\gamma_{\text{line}} = 2$. This is consistent with the framework described in Chapter 5.[6]

We add two uncontrollable renewable generators at nodes 4 and 5, so that $n_d = 14, n_g = 6$ and $n_w = 2$. The feasible space $\boldsymbol{\Theta} = \boldsymbol{\Theta}(\lambda)$ is a compact polytope that depends on λ, as shown in Fig. 7.1. All the calculations related to multiparametric programming are performed using the MPT3 toolbox [81].

We model the renewable generation as a 2-dimensional Gaussian random vector $\boldsymbol{\theta} \sim \mathcal{N}_2(\boldsymbol{\mu}_{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$, where $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ is interpreted as the nominal, or forecast, renewable generation, and $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ is computed based on the normalized symmetric graph Laplacian, following [91]:

$$\mathbf{C} = \tau^{2\kappa}(L_{\text{sym}} + \tau^2 I)^{-\kappa} \in \mathbb{R}^{n \times n}, \tag{7.35}$$

where L_{sym} is the normalized symmetric graph Laplacian

$$L_{\text{sym}} = \Delta^{-1/2} L_{\text{sym}} \Delta^{-1/2},$$

$\Delta \in \mathbb{R}^{n \times n}$ is the diagonal matrix with entries equal to the weighted nodal degrees $\Delta_{i,i} = \sum_{j \neq i} w_{i,j}$, and $\kappa, \tau^2 > 0$. The definition of the covariance matrix in Eq. (7.35) enables us to model positive correlations between neighboring (thus geographically "close") nodes.

In our experiments, we set $\kappa = 2$ and $\tau^2 = 1$. Then, we consider the $n_w \times n_w$ submatrix $\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}$ of $\mathbf{C}$ obtained by choosing rows and columns of $\mathbf{C}$ indexed by $\mathcal{N}_{n_w} = \{4, 5\}$, and we define $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ as

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}} := \operatorname{diag}(\{\delta_i\}_{i=1}^{n_w})\, \widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}\, \operatorname{diag}(\{\delta_i\}_{i=1}^{n_w}) \in \mathbb{R}^{n_w \times n_w}, \tag{7.36}$$

where the parameters δ_i's control the magnitudes of the standard deviations $\sigma_i := \sqrt{\boldsymbol{\Sigma}_{\boldsymbol{\theta}}(i,i)}$, $i = 1, 2$. In particular, the δ_i's are chosen in such a way that the standard deviations match realistic values for wind power forecasting error, expressed as a fraction of the corresponding installed capacity, over different time windows T (see also Chapter 2, Section 2.4.2):

$$\sigma_i = q(T) \times \mu_i^{(\text{installed})},\ i = 1, 2,$$

where $q = [0.01, 0.018, 0.04]$, corresponding to time windows of $5, 15$ and 60 minutes, respectively. Finally, the installed capacity of the renewable generators

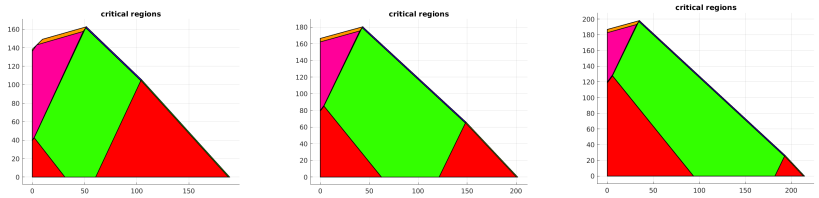
[6] The only difference is that, in Chapter 5, we had $\gamma_{\text{line}} = 1$, while here we may have to set $\gamma_{\text{line}} > 1$ in order to use a $\lambda < 1$. The reason is that, contrary to Chapter 5, we enforce generator limit constraints. As a consequence, it is possible that the OPF is not feasible if $\lambda < 1$.

are chosen based on the boundary of the 2-dimensional feasible space $\boldsymbol{\Theta}$, namely $\mu_1^{(\text{installed})} = \max\{x : (x,y) \in \boldsymbol{\Theta}\}, \mu_2^{(\text{installed})} = \max\{y : (x,y) \in \boldsymbol{\Theta}\}$.

Although $\boldsymbol{\theta} \sim \mathcal{N}_2(\boldsymbol{\mu}_{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$ is in principle unbounded, we choose the relevant parameters in such a way that, in practice, $\boldsymbol{\theta}$ never exceeds the boundary of the feasible space $\boldsymbol{\Theta}$. Note that the term ε is not present in the definition of $\boldsymbol{\theta}$, for the same reasons as explained in Section 4.1. Since $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ is obtained from realistic values for wind power forecasting error, the question is whether the matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ used in the numerics is close enough to the small-noise regime to make the large deviations results meaningful. As we show, the answer to this question is affirmative, validating the use of the large deviations methodology.

Exploratory analysis

We set $\gamma_{\text{line}} = 2$, and vary $\lambda \in \{0.6, 0.7, 0.8\}$. Fig. 7.1 shows the feasible space $\boldsymbol{\Theta}(\lambda)$, and its partition into critical regions, for different values of λ. Given



(a) $\lambda = 0.6$, 14 regions. (b) $\lambda = 0.7$, 13 regions. (c) $\lambda = 0.8$, 13 regions.

Figure 7.1: Partition of feasible space $\boldsymbol{\Theta}(\lambda)$ into critical regions for various values of λ.

$\boldsymbol{\mu}_{\boldsymbol{\theta}} \in \boldsymbol{\Theta}$, we set the price thresholds defining the spike event as

$$\boldsymbol{\alpha}_i^- = \text{LMP}_i(\boldsymbol{\mu}_{\boldsymbol{\theta}}) - \text{err}_{\text{rel}}|\text{LMP}_i(\boldsymbol{\mu}_{\boldsymbol{\theta}})|, \; \boldsymbol{\alpha}_i^+ = \text{LMP}_i(\boldsymbol{\mu}_{\boldsymbol{\theta}}) + \text{err}_{\text{rel}}|\text{LMP}_i(\boldsymbol{\mu}_{\boldsymbol{\theta}})|, \tag{7.37}$$

where $\text{err}_{\text{rel}} > 0$. In other words, we are interested in studying the event of a *relative price deviation* of magnitude greater than $\text{err}_{\text{rel}} > 0$:

$$\begin{aligned} Y &= \bigcup_{i=1}^{n} Y_i(p), \; Y_i(p) \\ &= \{\boldsymbol{\theta} \in \boldsymbol{\Theta} : |\text{LMP}_i(\boldsymbol{\theta}) - \text{LMP}_i(\boldsymbol{\mu}_{\boldsymbol{\theta}})| > \text{err}_{\text{rel}}|\text{LMP}_i(\boldsymbol{\mu}_{\boldsymbol{\theta}})|\}. \end{aligned}$$

Multimodality and sensitivity with respect to $\boldsymbol{\mu}_{\boldsymbol{\theta}}$

Next, we analyze in more detail the particular setting $\lambda = 0.6$, and consider three scenarios, corresponding to low, medium and high expected wind generation, i.e.

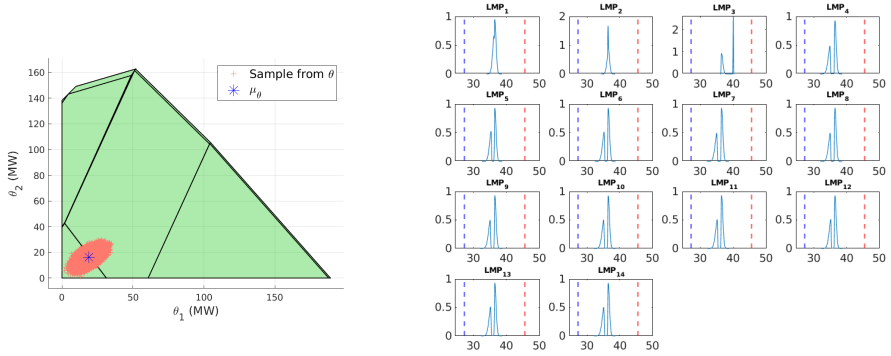
$\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(\text{low})} = 0.1\times\boldsymbol{\mu}^{(\text{installed})}$, $\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(\text{medium})} = 0.25\times\boldsymbol{\mu}^{(\text{medium})}$, $\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(\text{high})} = 0.5\times\boldsymbol{\mu}^{(\text{installed})}$, and $q^{(\text{medium})} = 0.018$. Fig. 7.2 shows the location of $\boldsymbol{\mu}_{\boldsymbol{\theta}}$, together with 10^6 samples from $\boldsymbol{\theta}$, and the corresponding empirical densities of the random variables LMP_i, $i = 1, \dots, 14$, obtained through Monte Carlo simulation. The red and blue vertical bars correspond to α_i^- and α_i^+ in (7.37) with a relative percentage error of $\text{err}_{\text{rel}} = 0.25$.

We observe that the results are extremely sensitive to the standard deviation and location (more so than the magnitude) of the forecast renewable generation $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ relative to the geometry of the critical regions, as this affects whether the samples of $\boldsymbol{\theta}$ will cross the boundary between adjacent regions or not. This confirms the importance of a holistic approach that combines distributional properties of the underlying random quantities with structural properties of the power grid operations.

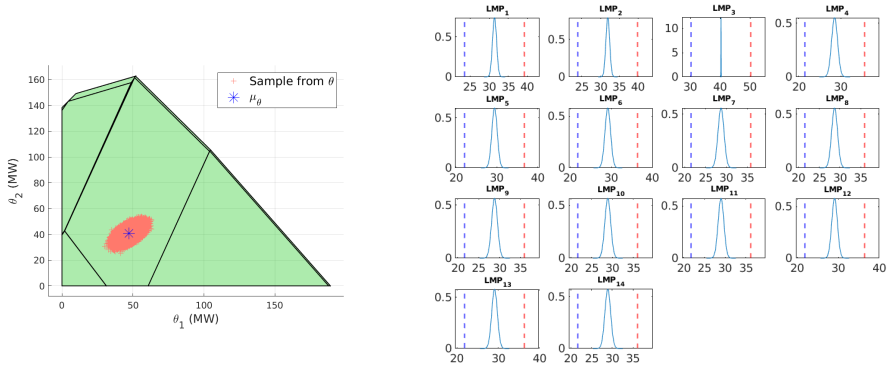
In turn, the crossing of a boundary can result in the distribution of the LMPs being *multimodal* (see Figs. 7.2a and 7.2c), due to the piecewise affine nature of the map $\boldsymbol{\theta} \to \mathbf{LMP}$. This observation shows how the problem of studying LMPs fluctuations is intrinsically harder than that of emergent line failures, as in [136, 137]. The phenomenon is more pronounced in the presence of steep gradient changes at the boundary between regions (or in the case of discontinuities), as can be observed in Fig. 7.3, whichs show the piecewise affine map $\boldsymbol{\theta} \to \mathbf{LMP}_{10}(\boldsymbol{\theta})$ for the three different choices of $\boldsymbol{\mu}_{\boldsymbol{\theta}}$. In particular, the expected $\mathbf{LMP}$ can differ greatly from $\mathbf{LMP}(\boldsymbol{\mu}_{\boldsymbol{\theta}})$.

7.5.1 Ranking of nodes based on their likelihood of having a price spike

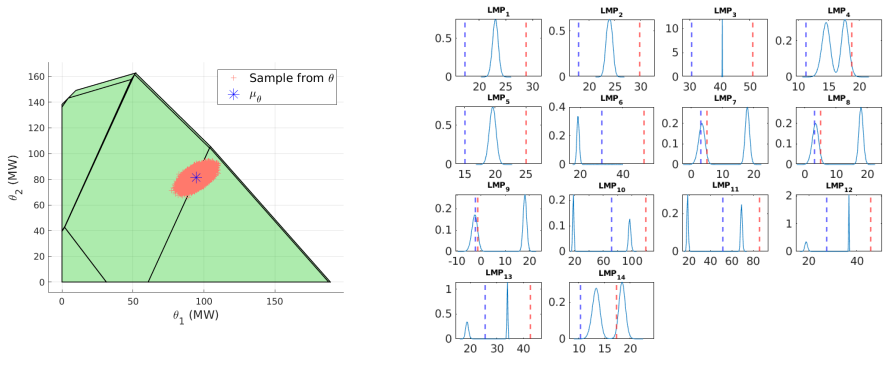
As illustrated by Eq. (7.29), large deviations theory predicts the most likely node to be $\arg\min_{i=1,\dots,n} I_i^*$, where $I_i^* := \inf_{\boldsymbol{\theta}\in\bigcup_{k=1}^M T_{i,k}} I(\boldsymbol{\theta})$. Indirectly, this approach produces also a *ranking* of nodes according to their likelihood of having a price spike. The use of large deviations theory to rank power grid components according to their likelihood of experiencing anomalous deviations from a nominal state has been validated in Chapter 4 in the context of transmission line failures. In order to validate the accuracy of the LDP methodology also for ranking nodes according to the likelihood of their price spikes, we compare the LD-based ranking with the one obtained via crude Monte Carlo simulation, as described in Table 7.1. We observe that the LD-based approach is able to recover the exact ranking of nodes, for various levels of relative error err_{rel}. Table 7.1 reports the values of the probability $\widehat{\mathbb{P}}(Y_i)$ of a price spike in node i, calculated using Monte Carlo simulation, together with the corresponding decay rates $I_i^* = \inf_{\boldsymbol{\theta}\in\bigcup_{k=1}^M T_{i,k}} I(\boldsymbol{\theta})$, showing that the LD-based approach correctly



(a) $\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(\text{low})} = 0.1 \times \boldsymbol{\mu}^{(\text{installed})}$.

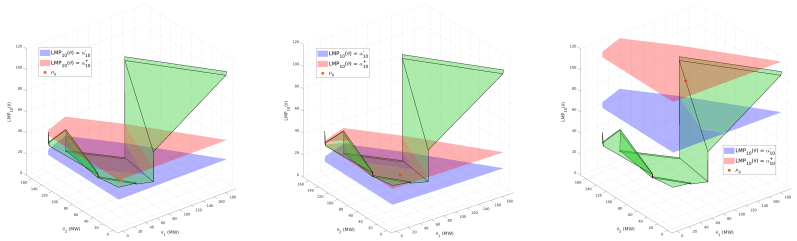


(b) $\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(\text{medium})} = 0.25 \times \boldsymbol{\mu}^{(\text{installed})}$.



(c) $\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(\text{high})} = 0.5 \times \boldsymbol{\mu}^{(\text{installed})}$.

Figure 7.2: Visualization of $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ and empirical distribution of $\boldsymbol{\theta}$ (left), and corresponding empirical densities of the random variables LMP_i, $i = 1, \ldots, 14$ (right), for different choices of $\boldsymbol{\mu}_{\boldsymbol{\theta}}$, $q = 0.018$.



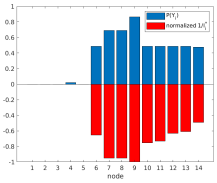
(a) $\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(\text{low})} = 0.1 \times \boldsymbol{\mu}^{(\text{installed})}$. (b) $\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(\text{medium})} = 0.25 \times \boldsymbol{\mu}^{(\text{installed})}$. (c) $\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(\text{high})} = 0.5 \times \boldsymbol{\mu}^{(\text{installed})}$.

Figure 7.3: Visualization of the piecewise affine map $\boldsymbol{\theta} \to \mathbf{LMP}_{10}(\boldsymbol{\theta})$ for three different locations of $\boldsymbol{\mu}_{\boldsymbol{\theta}}$, together with the price thresholds $\alpha_{10}^{\pm} = \text{LMP}_{10}(\boldsymbol{\mu}_{\boldsymbol{\theta}}) \pm \text{err}_{\text{rel}}|\text{LMP}_{10}(\boldsymbol{\mu}_{\boldsymbol{\theta}})|$.

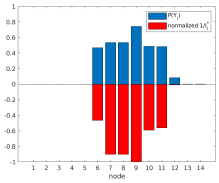
identifies the ranking. This property is validated more extensively in Fig. 7.4, which depicts the values of $\widehat{\mathbb{P}}(Y_i)$ against $-\min_k I_k^*/I_i^*$ across a wider range of price thresholds err_{rel}.

i	$\widehat{\mathbb{P}}(Y_i)$	I_i^*	rank
9	8.6371e-01	8.1160e-04	1
8	6.8984e-01	8.5572e-04	2
7	6.8984e-01	8.5572e-04	3
10	4.8713e-01	1.0786e-03	4
11	4.8690e-01	1.1123e-03	5
6	4.8613e-01	1.2438e-03	6
12	4.8586e-01	1.2849e-03	7
13	4.8586e-01	1.3296e-03	8
14	4.7559e-01	1.6548e-03	9
4	2.1282e-02	4.0854e+00	10
5	0	6.8384e+01	11
1	0	1.1584e+02	12
2	0	1.2971e+02	13
3	0	2.6984e+03	14

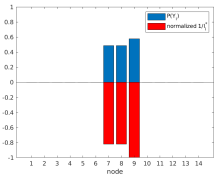
Table 7.1: Ranking of nodes based on the likelihood of having a price spike, according to both Monte Carlo simulation (in terms of probabilities $\widehat{\mathbb{P}}(Y_i)$) and large deviations results (in terms of decay rates I_i^*), for the case $\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(\text{high})} = 0.5 \times \boldsymbol{\mu}^{(\text{installed})}$, $\text{err}_{\text{rel}} = 0.25$, $q = 0.018$. The values $\widehat{\mathbb{P}}(Y_i)$, for $i = 1, 2, 3, 5$, are not reported as the Monte Carlo simulation is not sufficiently accurate for such small probabilities.



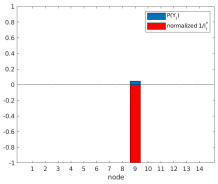
(a) $\text{err}_{\text{rel}} = 0.25$.



(b) $\text{err}_{\text{rel}} = 0.5$.



(c) $\text{err}_{\text{rel}} = 1$.



(d) $\text{err}_{\text{rel}} = 10$.

Figure 7.4: Comparison between empirical probabilities $\widehat{\mathbb{P}}(Y_i)$ based on Monte Carlo simulation and normalized decay rates $-\min_i I_i^*/I_i^*$ across a wider range of values for err_{rel}.

7.6 Concluding remarks and future work

In this chapter, we illustrate the potential of concepts from large deviations theory to study the events of rare price spikes caused by fluctuations of renewable generation. By assuming a centralized perspective, we use large deviations theory to approximate the probabilities of such events, and to rank the nodes of the power grids according to their likelihood of experiencing a price spike. Our technical approach is able to handle the multimodality of LMP's distributions, as well as violations of the LICQ regularity condition. Future research directions include extending the present framework to non-Gaussian fluctuations. Moreover, it would be of interest to study the sensitivity of the approximation in Eq. (7.28) with respect to the tuning parameter λ, which quantifies the conservatism in the choice of the line limits. This would allow us to establish a link with the analysis in Section 4.B.3, Chapter 4, and to extend the notion of safe capacity regions of Chapters 2 and 3 in the context of energy prices.

Bibliography

[1] R. Albert, I. Albert, and G. Nakarado. Structural vulnerability of the North American power grid. *Phys. Rev. E*, 69(2):025103, 2004.

[2] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74(1):47–97, 2002.

[3] R. Albert, H. Jeong, and A.-L. Barabási. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382, 2000.

[4] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. M. de Pison, and F. Antonanzas-Torres. Review of photovoltaic power forecasting. *Solar Energy*, 136:78 – 111, 2016.

[5] M. Babakmehr, M. G. Simões, M. B. Wakin, and F. Harirchi. Compressive sensing-based topology identification for smart grids. *IEEE Transactions on Industrial Informatics*, 12(2):532–543, 2016.

[6] X. Bai. Advance the ecosystem approach in cities. *Nature*, 559(7), 2018.

[7] P. Bak, C. Tang, and K. Wiesenfeld. Self-organized criticality. *Phys. Rev. A*, 38:364–374, 1988.

[8] R. B. Bapat. Resistance matrix and q-laplacian of a unicyclic graph. *Ramanujan Mathematical Society Lecture Notes Series*, 7:63–72, 2008.

[9] R. B. Bapat. *Graphs and matrices*. Springer, 2010.

[10] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[11] A. Bemporad, M. Morari, V. Dua, and E. N. Pistikopoulos. The explicit linear quadratic regulator for constrained systems. *Automatica*, 38(1):3 – 20, 2002.

[12] J. Berg, A. Natarajan, J. Mann, and E. Patton. Gaussian vs non-Gaussian turbulence: impact on wind turbine loads. *Wind Energy*, 19(11):1975–1989, 2016.

[13] A. Bernstein, D. Bienstock, D. Hay, M. Uzunoglu, and G. Zussman. Power grid vulnerability to geographically correlated failures — analysis and control implications. In *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, 2634–2642, 2014.

[14] D. Bienstock. *Electrical transmission system cascades and vulnerability - an operations research viewpoint*, volume 22 of *MOS-SIAM Series on Optimization*. SIAM, 2016.

[15] D. Bienstock, J. Blanchet, and J. Li. Stochastic models and control for electrical power line temperature. *Energy Systems*, 1–20, 2013.

[16] D. Bienstock, M. Chertkov, and S. Harnett. Chance-constrained optimal power flow: Risk-aware network control under uncertainty. *SIAM Review*, 56(3):461–495, 2014.

[17] D. Bienstock and A. Verma. Strong np-hardness of AC power flows feasibility. *Operations Research Letters*, 47(6):494 – 501, 2019.

[18] J. R. Birge, A. Hortaçsu, and J. M. Pavlin. Inverse optimization for the recovery of market structure from market outcomes: An application to the miso electricity market. *Operations Research*, 65(4):837–855, 2017.

[19] R. Bo and F. Li. Probabilistic LMP forecasting considering load uncertainty. *IEEE Transactions on Power Systems*, 24(3):1279–1289, 2009.

[20] R. Bo and F. Li. Probabilistic LMP forecasting under AC optimal power flow framework: Theory and applications. *Electric Power Systems Research*, 88(Supplement C):16 – 24, 2012.

[21] S. Bolognani and F. Dörfler. Fast scenario-based decision making in unbalanced distribution networks. In *2016 Power Systems Computation Conference (PSCC)*, 1–7, 2016.

[22] J. F. Bonnans and A. Shapiro. *Perturbation analysis of optimization problems*. Springer, 2000.

[23] A. S. Brouwer, M. van den Broek, A. Seebregts, and A. Faaij. Impacts of large-scale intermittent renewable energy sources on electricity systems, and how these can be modeled. *Renewable and Sustainable Energy Reviews*, 33(Supplement C):443 – 466, 2014.

[24] T. Brown. https://pypsa.org/examples/scigrid-lopf-then-pf.html, 2017.

[25] T. Brown. https://pypsa.org/examples/add_load_gen_trafos_to_scigrid.html, 2017.

[26] T. Brown, J. Hörsch, and D. Schlachtberger. Pypsa: Python for power system analysis. *Journal of Open Research Software*, 6(1), 2018.

[27] J. A. Bucklew. *Large deviation techniques in decision, simulation, and estimation*, volume 190. Wiley New York, 1990.

[28] H. N. Byström. Extreme value theory and extremely large electricity price changes. *International Review of Economics & Finance*, 14(1):41–55, 2005.

[29] M. Cain, R. O'Neill, and A. Castillo. History of optimal power flow and formulations. Technical report, US Federal Energy Regulatory Commission, 2012.

[30] G. Calafiore and M. C. Campi. Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming*, 102(1):25–46, 2005.

[31] G. C. Calafiore and M. C. Campi. The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 51(5):742–753, 2006.

[32] California Independent System Operator. https://www.caiso.com.

[33] California ISO. Appendix c: Locational marginal price. https://www.caiso.com/Documents/AppendicesC-F-FifthReplacementCAISOTariff_15-Dec-10.pdf.

[34] California ISO. Managing oversupply. http://www.caiso.com/informed/Pages/ManagingOversupply.aspx.

[35] California ISO. Market price maps. http://www.caiso.com/PriceMap/Pages/default.aspx.

[36] California ISO. Using renewables to operate a low-carbon grid. https://www.caiso.com/Documents/UsingRenewablesToOperateLowCarbonGrid-FAQ.pdf.

[37] S. Carmi, Z. Wu, S. Havlin, and H. Stanley. Transport in networks with multiple sources and sinks. *Europhysics Letters*, 84(2):28005, 2008.

[38] B. A. Carreras, V. E. Lynch, I. Dobson, and D. E. Newman. Complex dynamics of blackouts in power transmission systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 14(3):643–652, 2004.

[39] B. A. Carreras, D. E. Newman, and I. Dobson. North american blackout time series statistics and implications for blackout risk. *IEEE Transactions on Power Systems*, 31(6):4406–4414, 2016.

[40] H. Cetinay, F. A. Kuipers, and P. Van Mieghem. A topological investigation of power flow. *IEEE Systems Journal*, 12(3):2524–2532, 2018.

[41] M. Chertkov, F. Pan, and M. Stepanov. Predicting Failures in Power Grids: The Case of Static Overloads. *IEEE Transactions on Smart Grid*, 2(1):162–172, 2011.

[42] A. Clauset, C. R. Shalizi, and M. E. Newman. Power-law distributions in empirical data. *SIAM review*, 51(4):661–703, 2009.

[43] C. Coffrin, P. Van Hentenryck, and R. Bent. Approximating line losses and apparent power in AC power flow linearizations. In *2012 IEEE Power and Energy Society General Meeting*, 1–8, 2012.

[44] R. Cohen, K. Erez, D. Ben-Avraham, and S. Havlin. Resilience of the Internet to random breakdowns. *Phys. Rev. Lett.*, 85(21):4626–4628, 2000.

[45] R. Cohen, K. Erez, D. Ben-Avraham, and S. Havlin. Breakdown of the Internet under Intentional Attack. *Phys. Rev. Lett.*, 86(16):3682–3685, 2001.

[46] G. Constable and B. Somerville. *A century of innovation: Twenty engineering achievements that transformed our lives.* Joseph Henry Press, 2003.

[47] P. Cramton. Electricity market design. *Oxford Review of Economic Policy*, 33(4):589–612, 2017.

[48] P. Crucitti, V. Latora, and M. Marchiori. A topological analysis of the Italian electric power grid. *Physica A: Statistical Mechanics and its Applications*, 338(1-2):92–97, 2004.

[49] P. Crucitti, V. Latora, and M. Marchiori. Model for cascading failures in complex networks. *Phys. Rev. E*, 69(4):045104, 2004.

[50] P. Crucitti, V. Latora, M. Marchiori, and A. Rapisarda. Efficiency of scale-free networks: error and attack tolerance. *Physica A: Statistical Mechanics and its Applications*, 320:622–642, 2003.

[51] J. P. Cunningham, P. Hennig, and S. Lacoste-Julien. Gaussian probabilities and expectation propagation. *ArXiv e-prints:* 1111.6832, 2011.

[52] G. Dasarathy. A simple probability trick for bounding the expected maximum of n random variables. http://www.public.asu.edu/~gdasarat/files/maxGaussians.pdf, 2011.

[53] Data Packaged Core Datasets. Nuts_rg_60m_2013.shp. https://github.com/datasets/geo-nuts-administrative-boundaries/blob/master/data/NUTS_2013_60M_SH/data/NUTS_RG_60M_2013.shp. [Accessed on October 2018].

[54] G. S. de Miera, P. del Río González, and I. Vizcaíno. Analysing the impact of renewable electricity support schemes on power prices: The case of wind electricity in spain. *Energy Policy*, 36(9):3345 – 3359, 2008.

[55] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*. Springer, 1998.

[56] D. P. Dobkin, A. R. Wilks, S. V. Levy, and W. P. Thurston. Contour tracing by piecewise linear approximations. *ACM Transactions on Graphics (TOG)*, 9(4):389–423, 1990.

[57] I. Dobson, B. Carreras, V. Lynch, and D. Newman. Complex systems analysis of series of blackouts: Cascading failure, critical points, and self-organization. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 17(2):026103, 2007.

[58] F. Dörfler, M. Chertkov, and F. Bullo. Synchronization in complex oscillator networks and smart grids. *Proceedings of the National Academy of Sciences*, 110(6):2005–2010, 2013.

[59] C. Eid, P. Codani, Y. Perez, J. Reneses, and R. Hakvoort. Managing electric flexibility from distributed energy resources: A review of incentives for market design. *Renewable and Sustainable Energy Reviews*, 64:237 – 247, 2016.

[60] Energy Information Administration. Fewer wind curtailments and negative power prices seen in texas after major grid expansion. https://www.eia.gov/todayinenergy/detail.php?id=16831.

[61] Eurostat. http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=nama_10r_3popgdp. [Accessed on October 2018].

[62] Eurostat. Nuts_rg_60m_2013.shp. https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts#nuts13. [Accessed on October 2018].

[63] Federal Energy Regulatory Commission, North American Electric Reliability Corporation. Arizona-southern california outages on september 8, 2011: Causes and recommendations. Technical report, 2012.

[64] FERC. White paper wholesale power market platform. 2003.

[65] Florence School of Regulation. Zonal versus nodal electricity pricing: the pjm experience. https://fsr.eui.eu/zonal-versus-nodal-electricity-pricing-the-pjm-experience/.

[66] J. H. Friedman. Multivariate adaptive regression splines. *The Annals of Statistics*, 19(1):1–67, 1991.

[67] Frontier Economics. International transmission pricing review. https://www.ea.govt.nz/dmsdocument/2539.

[68] A. J. Ganesh, N. O'Connell, and D. J. Wischik. *Big queues*. Springer, 2004.

[69] J. Gao, B. Barzel, and A.-L. Barabási. Universal resilience patterns in complex networks. *Nature*, 530(7590):307, 2016.

[70] X. Geng and L. Xie. Learning the LMP-load coupling from data: A support vector machine based approach. *IEEE Transactions on Power Systems*, 32(2):1127–1138, 2017.

[71] A. Gerster. Negative price spikes at power markets: the role of energy policy. *Journal of Regulatory Economics*, 50(3):271-289, 2016.

[72] H. Glavitsch and R. Bacher. Optimal power flow algorithms. In *Analysis and Control System Techniques for Electric Power Systems, Part 1 of 4*, volume 41 of *Control and Dynamic Systems*, 135 – 205. 1991.

[73] K.-I. Goh, D.-S. Lee, B. Kahng, and D. Kim. Sandpile on scale-free networks. *Phys. Rev. Lett.*, 91:148701, 2003.

[74] A. Gomez-Exposito, A. J. Conejo, and C. Canizares. *Electric energy systems: analysis and operation*, CRC press.

[75] J. Gonzalez, J. Moriarty, and J. Palczewski. Bayesian calibration and number of jump components in electricity spot price models. *Energy Economics,* 65:375-388, 2017.

[76] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. http://cvxr.com/cvx, 2014.

[77] J. Guo, Y. Fu, Z. Li, and M. Shahidehpour. Direct calculation of line outage distribution factors. *IEEE Transactions on Power Systems*, 24(3):1633–1634, 2009.

[78] L. I. Hagfors, H. H. Kamperud, F. Paraschiv, M. Prokopczuk, A. Sator, and S. Westgaard. Prediction of extreme price occurrences in the german day-ahead electricity market. *Quantitative Finance*, 16(12):1929–1948, 2016.

[79] K. W. Hedman, S. S. Oren and R. P. O'Neill. A review of transmission switching and network topology optimization. *2011 IEEE Power and Energy Society General Meeting*, 1-7, 2011.

[80] D. Heide, M. Schäfer, and M. Greiner. Robustness of networks against fluctuation-induced cascading failures. *Phys. Rev. E*, 77(5):056103, 2008.

[81] M. Herceg, M. Kvasnica, C. Jones, and M. Morari. Multi-Parametric Toolbox 3.0. In *Proc. of the European Control Conference*, 502–510, 2013. http://control.ee.ethz.ch/~mpt.

[82] M. Heymann and E. Vanden-Eijnden. The geometric minimum action method: A least action principle on the space of curves. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 61(8):1052–1117, 2008.

[83] H. Higgs and A. Worthington. Stochastic price modeling of high volatility, mean-reverting, spike-prone commodities: The Australian wholesale spot electricity market. *Energy Economics*, 30(6):3172–3185, 2008.

[84] B. M. Hill. A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, 3(5):1163–1174, 1975.

[85] J. Hindes and I. B. Schwartz. Epidemic extinction and control in heterogeneous networks. *Phys. Rev. Lett.*, 117:028302, 2016.

[86] P. Hines, K. Balasubramaniam, and E. C. Sanchez. Cascading failures in power grids. *IEEE Potentials*, 28(5):24–30, 2009.

[87] P. Hines, I. Dobson, E. Cotilla-Sanchez, and M. Eppstein. "Dual Graph" and "Random Chemistry" Methods for Cascading Failure Analysis. In *2013 46th Hawaii International Conference on System Sciences*, 2141–2150, 2013.

[88] P. D. H. Hines, I. Dobson, and P. Rezaei. Cascading power outages propagate locally in an influence graph that is not the actual grid topology. *IEEE Transactions on Power Systems*, 32(2):958–967, 2017.

[89] B. Hodge and M. Milligan. Wind power forecasting error distributions over multiple timescales. In *2011 IEEE Power and Energy Society General Meeting*, 1–8, 2011.

[90] B. Hodge, S. Shedd, and A. Florita. Examining the variability of wind power output in the regulation time frame. Technical report, National Renewable Energy Laboratory, 2012.

[91] F. Hoffmann, B. Hosseini, Z. Ren, and A. M. Stuart. Consistency of semi-supervised learning algorithms on graphs: Probit and one-hot methods. *ArXiv e-prints:* 1906.07658, 2019.

[92] L. Huang, L. Yang, and K. Yang. Geographical effects on cascading breakdowns of scale-free networks. *Phys. Rev. E*, 73:036102, 2006.

[93] R. Huang, T. Huang, R. Gadh, and N. Li. Solar generation prediction using the arma model in a laboratory-level micro-grid. In *2012 IEEE Third International Conference on Smart Grid Communications (SmartGridComm)*, 528–533, 2012.

[94] M. Huneault and F. D. Galiana. A survey of the optimal power flow literature. *IEEE Transactions on Power Systems*, 6(2):762–770, 1991.

[95] E. B. Iversen, J. M. Morales, J. K. Møller, and H. Madsen. Short-term probabilistic forecasting of wind speed using stochastic differential equations. *International Journal of Forecasting*, 32(3):981 – 990, 2016.

[96] E. B. Iversen, J. M. Morales, J. K. Møller, P.-J. Trombe, and H. Madsen. Leveraging stochastic differential equations for probabilistic forecasting of wind power using a dynamic power curve. *Wind Energy*, 20(1):33–44, 2017.

[97] Y. Ji, R. J. Thomas, and L. Tong. Probabilistic forecasting of real-time LMP and network congestion. *IEEE Transactions on Power Systems*, 32(2):831–841, 2017.

[98] I.T. Jolliffe. Principal Component Analysis. *Springer Series in Statistics*, 2002.

[99] D. Jung and S. Kettemann. Long-range response in AC electricity grids. *Phys. Rev. E*, 94:012307, 2016.

[100] V. Kaminski. *Energy markets*. Risk Books, 2012.

[101] V. Kekatos, G. B. Giannakis, and R. Baldick. Online energy price matrix factorization for power grid topology tracking. *IEEE Transactions on Smart Grid*, 7(3):1239–1248, 2016.

[102] J. Kersulis, I. Hiskens, M. Chertkov, S. Backhaus, and D. Bienstock. Temperature-based instanton analysis: Identifying vulnerability in transmission networks. In *2015 IEEE Eindhoven PowerTech*, 1–6. IEEE, 2015.

[103] S. Kettemann. Delocalization of disturbances and the stability of AC electricity grids. *Phys. Rev. E*, 94:062311, 2016.

[104] J. Kim and I. Dobson. Approximating a loading-dependent cascading failure model with a branching process. *IEEE Transactions on Reliability*, 59(4):691–699, 2010.

[105] R. Kinney, P. Crucitti, R. Albert, and V. Latora. Modeling cascading failures in the North American power grid. *European Physical Journal B*, 46(1):101–107, 2005.

[106] S. Kolumban, S. Kapodistria, and N. Nooraee. Short and long-term wind turbine power output prediction. *ArXiv e-prints:* 1707.06497, 2017.

[107] M. Korkali, J. G. Veneman, B. F. Tivnan, J. P. Bagrow, and P. D. Hines. Reducing cascading failure risk by increasing infrastructure network interdependence. *Scientific reports*, 7:44499, 2017.

[108] B. Kroposki, B. Johnson, Y. Zhang, V. Gevorgian, P. Denholm, B. Hodge, and B. Hannegan. Achieving a 100% renewable grid: Operating electric power systems with extremely high levels of variable renewable energy. *IEEE Power and Energy Magazine*, 15(2):61–73, 2017.

[109] D. Labavić, R. Suciu, H. Meyer-Ortmanns, and S. Kettemann. Long-range response to transmission line disturbances in dc electricity grids. *The European Physical Journal Special Topics*, 223(12):2517–2525, 2014.

[110] P. L'Ecuyer, V. Demers, and B. Tuffin. Splitting for rare-event simulation. *Proceedings of the 2006 Winter Simulation Conference*, 137–148, 2006.

[111] F. Li and R. Bo. Congestion and price prediction under load variation. *IEEE Transactions on Power Systems*, 24(2):911–922, 2009.

[112] Liu, H., Tesfatsion, L. and Chowdhury, A. A. Derivation of locational marginal prices for restructured wholesale power markets. *Journal of Energy Markets*, 2(1), 3, 2009.

[113] X. Lu, Z. Y. Dong, and X. Li. Electricity market price spike forecast with data mining techniques. *Electric Power Systems Research*, 73(1):19 – 29, 2005.

[114] J. MacQueen. Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 1(14), 1967.

[115] A. Malvaldi, S. Weiss, D. Infield, J. Browell, P. Leahy, and A. M. Foley. A spatial and temporal correlation analysis of aggregate wind power in an ideally interconnected europe. *Wind Energy*, 20(8):1315–1329, 2017.

[116] D. Manik, M. Rohden, H. Ronellenfitsch, X. Zhang, S. Hallerberg, D. Witthaut, and M. Timme. Network susceptibilities: Theory and applications. *Phys. Rev. E*, 95:012319, 2017.

[117] A. Mas-Colell, M. Whinston, and J. Green. *Microeconomic theory.* New York: Oxford University Press, 1995

[118] C. Matke, W. Medjroubi, and D. Kleinhans. SciGRID - An Open Source Reference Model for the European Transmission Network. http://scigrid.de, 2015.

[119] J. McCalley. The DC power flow equations. http:/home.engineering.iastate.edu/~jdm/ee553/DCPowerFlowEquations.pdf, 2012.

[120] D. Mehta, D. Molzahn, and K. Turitsyn. Recent advances in computational methods for the power flow equations. In *2016 American Control Conference (ACC)*, 1753–1765, 2016.

[121] P. Milan, M. Wächter, and J. Peinke. Turbulent character of wind energy. *Phys. Rev. Lett.*, 110:138701, 2013.

[122] M. Milligan, M. Schwartz, and Y. Wan. Statistical wind power forecasting models: Results for us wind farms. *National Renewable Energy Laboratory*, 2003.

[123] A. Mills and R. Wiser. Implications of wide-area geographic diversity for short-term variability of solar power. Technical report, Lawrence Berkeley National Laboratory, 2010.

[124] B. Mirzasoleiman, M. Babaei, M. Jalili, and M. Safari. Cascaded failures in weighted networks. *Phys. Rev. E*, 84(4):046114, 2011.

[125] J. K. Møller, M. Zugno, and H. Madsen. Probabilistic forecasts of wind power generation by stochastic differential equation models. *Journal of Forecasting*, 35(3):189–205, 2016.

[126] D. K. Molzahn, B. C. Lesieutre, and C. L. DeMarco. A sufficient condition for power flow insolvability with applications to voltage stability margins. *IEEE Transactions on Power Systems*, 28(3):2592–2601, 2013.

[127] F. Morone and H. A. Makse. Influence maximization in complex networks through optimal percolation. *Nature*, 524(7563):65, 2015.

[128] A. Motter. Cascade control and defense in complex networks. *Phys. Rev. Lett.*, 93(9):1–4, 2004.

[129] A. Motter and Y.-C. Lai. Cascade-based attacks on complex networks. *Phys. Rev. E*, 66(6):065102, 2002.

[130] C. E. Murillo-Sánchez, R. D. Zimmerman, C. L. Anderson, and R. J. Thomas. Secure planning and operations of systems with stochastic sources, energy storage, and active demand. *IEEE Transactions on Smart Grid*, 4(4):2220–2229, 2013.

[131] National Academies of Sciences, Engineering, and Medicine. *Analytic Research Foundations for the Next-Generation Electric Grid.* The National Academies Press, 2016.

[132] A. Nemirovski and A. Shapiro. Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17(4):969–996, 2006.

[133] T. Nesti, J. Nair, and B. Zwart. Temperature overloads in power grids under uncertainty: A large deviations approach. *IEEE Transactions on Control of Network Systems*, 6(3):1161–1173, 2019.

[134] T. Nesti, F. Sloothaak, and B. Zwart. Emergence of scale-free blackout sizes in power grids. In preparation.

[135] T. Nesti, J. Moriarty,A. Zocca, and B.Zwart. Large Fluctuations in Locational Marginal Prices. *ArXiv e-prints:* 2002.11680, 2020.

[136] T. Nesti, A. Zocca, and B. Zwart. Line failure probability bounds for power grids. In *2017 IEEE Power Energy Society General Meeting*, 1–5, 2017.

[137] T. Nesti, A. Zocca, and B. Zwart. Emergent failures and cascades in power grids: A statistical physics perspective. *Phys. Rev. Lett.*, 120:258301, 2018.

[138] K. Neuhoff, B. F. Hobbs, and D. Newbery. Congestion Management in European Power Networks: Criteria to Assess the Available Options. Discussion Papers of DIW Berlin 1161, German Institute for Economic Research, 2011.

[139] OpenStreetMap contributors. Planet dump retrieved from https://planet.osm.org. https://www.openstreetmap.org, 2017.

[140] T. Orfanogianni and G. Gross. A general formulation for LMP evaluation. *IEEE Transactions on Power Systems*, 22(3):1163–1173, 2007.

[141] F. Paraschiv, D. Erni, and R. Pietsch. The impact of renewable energies on eex day-ahead electricity prices. *Energy Policy*, 73:196 – 210, 2014.

[142] F. Paraschiv, S.-E. Fleten, and M. Schürle. A spot-forward model for electricity prices with regime shifts. *Energy Economics*, 47:142–153, 2015.

[143] F. Paraschiv, R. Hadzi-Mishev, and D. Keles. Extreme value theory for heavy tails in electricity prices. *Journal of Energy Markets*, 9(2), 2016.

[144] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, (14):3200, 2001.

[145] I. Pavić, M. Beus, H. Pandžić, T. Capuder, and I. Štritof. Electricity markets overview — market participation possibilities for renewable and distributed energy resources. In *2017 14th International Conference on the European Energy Market (EEM)*, 1–5, 2017.

[146] Y. Peings and G. Magnusdottir. Forcing of the wintertime atmospheric circulation by the multidecadal fluctuations of the north atlantic ocean. *Environmental Research Letters*, 9(3):034018, 2014.

[147] H. Pender and W. Del Mar. *Electrical engineers' handbook*. Wiley, 1949.

[148] T. Pesch, H.-J. Allelein, and J.-F. Hake. Impacts of the transformation of the German energy system on the transmission grid. *The European Physical Journal Special Topics*, 223(12):2561–2575, 2014.

[149] R. Plemmons. M-matrix characterizations. I–nonsingular m-matrices. *Linear Algebra and its Applications*, 18(2):175 – 188, 1977.

[150] L. Powell. *Power system load flow analysis*. McGraw Hill Professional, 2004.

[151] K. Purchala, L. Meeus, D. Van Dommelen, and R. Belmans. Usefulness of DC power flow for active power flow analysis. In *IEEE Power Engineering Society General Meeting, 2005*, 454–459, 2005.

[152] J. Qi, K. Sun, and S. Mei. An Interaction Model for Simulation and Mitigation of Cascading Failures. *IEEE Transactions on Power Systems*, 30(2):804–819, 2015.

[153] F. Radicchi and S. Fortunato. Explosive percolation in scale-free networks. *Phys. Rev. Lett.*, 103:168701, 2009.

[154] A. Radovanovic, T. Nesti, and B. Chen. A holistic approach to forecasting wholesale energy market prices. *IEEE Transactions on Power Systems*, 34(6):4317–4328, 2019.

[155] REN21. Renewables 2016 global status report. Technical report, Ren21 Secretariat, 2019.

[156] S. I. Resnick. *Heavy-tail phenomena*. Springer Series in Operations Research and Financial Engineering. 2007.

[157] L. Roald and G. Andersson. Chance-constrained AC optimal power flow: Reformulations and efficient algorithms. *IEEE Transactions on Power Systems*, 33(3):2906–2918, 2018.

[158] L. Roald, G. Andersson, S. Misra, M. Chertkov, and S. Backhaus. Optimal power flow with wind power control and limited expected risk of overloads. In *2016 Power Systems Computation Conference (PSCC)*, 1–7, 2016.

[159] L. Roald, S. Misra, M. Chertkov, and G. Andersson. Optimal power flow with weighted chance constraints and general policies for generation control. In *2015 54th IEEE Conference on Decision and Control (CDC)*, 6927–6933, 2015.

[160] H. Ronellenfitsch, D. Manik, J. Horsch, T. Brown, and D. Witthaut. Dual theory of transmission line outages. *IEEE Transactions on Power Systems*, 32(5):4060–4068, 2017.

[161] G. Rubino and B. Tuffin. *Rare Event Simulation Using Monte Carlo Methods*. Wiley Publishing, 2009.

[162] B. Schäfer, C. Beck, K. Aihara, D. Witthaut, and M. Timme. Non-Gaussian power grid frequency fluctuations characterized by lévy-stable laws and superstatistics. *Nature Energy*, 3(2):119, 2018.

[163] B. Schäfer, D. Witthaut, M. Timme, and V. Latora. Dynamically induced cascading failures in supply networks. *ArXiv e-prints:* 1707.08018, 2017.

[164] B. Schäfer, D. Witthaut, M. Timme, and V. Latora. Dynamically induced cascading failures in power grids. *Nature communications*, 9(1):1975, 2018.

[165] M. Schaub, J. Lehmann, S. Yaliraki, and M. Barahona. Structure of complex networks: Quantifying edge-to-edge relations by failure-induced flow redistribution. *Network Science*, 2(01):66–89, 2014.

[166] D. Schlachtberger, S. Becker, S. Schramm, and M. Greiner. Backup flexibility classes in emerging large-scale renewable electricity systems. *Energy Conversion and Management*, 125(Supplement C):336 – 346, 2016.

[167] J. F. Shortle. Efficient simulation of blackout probabilities using splitting. *International Journal of Electrical Power & Energy Systems*, 44(1):743 – 751, 2013.

[168] I. Simonsen, L. Buzna, K. Peters, S. Bornholdt, and D. Helbing. Transient dynamics increasing network vulnerability to cascading failures. *Phys. Rev. Lett.*, 100:218701, 2008.

[169] J. W. Simpson-Porco, F. Dörfler, and F. Bullo. Voltage collapse in complex power grids. *Nature Communications*, 7(10790), 2016.

[170] F. Sloothaak. Criticality in power networks: a probabilistic approach. Accepted/In Press, 2020.

[171] S. Soltan, D. Mazauric, and G. Zussman. Analysis of Failures in Power Grids. *IEEE Transactions on Control of Network Systems*, 4(2):288–300, 2017.

[172] Southwest Power Pool. 2016 annual state of the market report. https://www.spp.org/documents/53549/spp_mmu_asom_2016.pdf.

[173] S&P Global. https://www.spglobal.com/marketintelligence/en/campaigns/energy.

[174] S&P Global. California's renewable power saga is just beginning. https://blogs.platts.com/2015/06/18/california-renewable-power-saga/.

[175] B. Stott, J. Jardim, and O. Alsac. Dc power flow revisited. *IEEE Transactions on Power Systems*, 24(3):1290–1300, 2009.

[176] T. Summers, J. Warrington, M. Morari, and J. Lygeros. Stochastic optimal power flow based on convex approximations of chance constraints. In *2014 Power Systems Computation Conference*, 1–7, 2014.

[177] J. Sun and L. Tesfatsion. DC optimal power flow formulation and solution using quadprogj. 2010.

[178] K. Sun, Y. Hou, W. Sun, and J. Qi. *Power system control under cascading failures: understanding, mitigation, and system restoration.* Wiley-Blackwell, 2018.

[179] S. Sun, Z. Liu, Z. Chen, and Z. Yuan. Error and attack tolerance of evolving networks with local preferential attachment. *Physica A: Statistical Mechanics and its Applications*, 373:851–860, 2007.

[180] W. Tang and R. Jain. A nash equilibrium need not exist in the locational marginal pricing mechanism. *ArXiv e-prints:* 1310.4282, 2013.

[181] P. TøNdel, T. A. Johansen, and A. Bemporad. An algorithm for multiparametric quadratic programming and explicit mpc solutions. *Automatica*, 39(3):489–497, 2003.

[182] H. Touchette. The large deviation approach to statistical mechanics. *Physics Reports*, 478(1):1 – 69, 2009.

[183] U.S. - Canada Power System Outage Task Force. Final report on the August 14, 2003 blackout in the United States and Canada: Causes and recommendations. Technical report, 2004.

[184] US Department of Energy. Electric emergency incident and disturbance report (form OE-417). https://www.oe.netl.doe.gov/OE417_annual_summary.aspx. [Accessed November 2018].

[185] P. Van Mieghem. *Graph Spectra for Complex Networks.* Cambridge University Press, 2010.

[186] P. Van Mieghem, K. Devriendt, and H. Cetinay. Pseudoinverse of the laplacian and best spreader node in a network. *Phys. Rev. E*, 96:032311, 2017.

[187] A. E. D. Veraart. Modelling the impact of wind power production on electricity prices by regime-switching lévy semistationary processes. In *Stochastics of Environmental and Financial Economics*. Springer International Publishing, 2016.

[188] M. Vrakopoulou, M. Katsampani, K. Margellos, J. Lygeros, and G. Andersson. Probabilistic security-constrained AC optimal power flow. In *2013 IEEE Grenoble Conference*, 1–6, 2013.

[189] M. Vrakopoulou, K. Margellos, J. Lygeros, and G. Andersson. *Probabilistic Guarantees for the N-1 Security of Systems with Wind Power Generation*, 59–73. Springer India, 2013.

[190] W. Wadman. *Assessing power grid reliability using rare event simulation*. Universiteit van Amsterdam, 2015.

[191] W. Wadman, G. Bloemhof, D. Crommelin, and J. Frank. Probabilistic power flow simulation allowing temporary current overloading. In *Proceedings of the International Conference on Probabilistic Methods Applied to Power Systems*, 2012.

[192] W. S. Wadman, D. T. Crommelin, and B. P. Zwart. A large deviation based splitting estimation of power flow reliability. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 26(4):23, 2016.

[193] M. J. Wainwright. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2019.

[194] H. Wan, J. McCalley, and V. Vittal. Increasing thermal rating by risk analysis. *IEEE Transactions on Power Systems*, 14(3):815–828, 1999.

[195] Z. Wang, A. Scaglione, and R. J. Thomas. Generating statistically correct random topologies for testing smart grid communication and control networks. *IEEE Transactions on Smart Grid*, 1(1):28–39, 2010.

[196] D. Watts. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, 99(9):5766–5771, 2002.

[197] Weihui Fu and J. D. McCalley. Risk based optimal power flow. In *2001 IEEE Porto Power Tech Proceedings (Cat. No.01EX502)*, volume 3, page 6, 2001.

[198] R. Weron. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4):1030 – 1081, 2014.

[199] D. Witthaut, M. Rohden, X. Zhang, S. Hallerberg, and M. Timme. Critical Links and Nonlocal Rerouting in Complex Supply Networks. *Phys. Rev. Lett.*, 116(13):138701, 2016.

[200] D. Witthaut and M. Timme. Nonlocal failures in complex supply networks by single link additions. *The European Physical Journal B*, 86(9):377, 2013.

[201] D. Witthaut and M. Timme. Nonlocal effects and countermeasures in cascading failures. *Phys. Rev. E*, 92(3):032809, 2015.

[202] A. Wood, B. Wollenberg, and G. Sheble. *Power generation, operation, and control.* John Wiley & Sons, 3rd edition, 2014.

[203] World Population Review. http://worldpopulationreview.com/countries/germany-population/cities/. [Accessed on October 2018].

[204] Y. Yang. Hybrid grids. Towards a Hybrid AC/DC Transmission Grid. Technical report, DNV GL Strategic research innovation position paper 2, 2015.

[205] Y. Yang and A. E. Motter. Cascading failures as continuous phase-space transitions. *Phys. Rev. Lett.*, 119:248302, 2017.

[206] Y. Yang, T. Nishikawa, and A. E. Motter. Small vulnerable sets determine large network cascades in power grids. *Science*, 358(6365), 2017.

[207] Y. Yang, T. Nishikawa, and A. E. Motter. Vulnerability and cosusceptibility determine the size of network cascades. *Phys. Rev. Lett.*, 118:048301, 2017.

[208] J. Zhang, B.-M. Hodge, and A. Florita. Investigating the Correlation Between Wind and Solar Power Forecast Errors in the Western Interconnection. ASME 2013 7th International Conference on Energy Sustainability of *Energy Sustainability*, 2013. V001T16A003.

[209] Q. Zhou, L. Tesfatsion, and C. C. Liu. Short-term congestion forecasting in wholesale power markets. *IEEE Transactions on Power Systems*, 26(4):2185–2196, 2011.

[210] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas. Matpower: Steady-state operations, planning, and analysis tools for power systems research and education. *Power Systems, IEEE Transactions on*, 26(1):12–19, 2011.

Summary

In this thesis, we develop probabilistic techniques for the analysis of energy systems under uncertainty. The research is motivated by the advent of renewable energy sources, such as wind and solar photovoltaics, which brought about a paradigm shift for the design and control of power grids. To ensure that power grids reliably and steadily deliver power to customers, stability constraints such as admissible ranges on power flows must be satisfied at all times. The inherently uncertain nature of renewable energy sources is responsible for considerable supply-side variability and may lead to unexpected violations of stability constraints, which can cause the failure of grid components and result in widespread blackouts with enormous societal costs. Electricity prices are also affected by the increased uncertainty in power outputs, resulting in highly variable prices that can even turn negative during periods of low demand and high renewable production. This thesis introduces powerful tools such as large deviations theory, concentration inequalities, and statistical learning techniques, that can guide the uncertainty-aware operations of power grids. Chapter 1 provides an overview of the power grid operations landscape, presents the relevant literature, and illustrates our main contributions to the field.

Chapters 2 and 3 focus on developing probabilistic counterparts of traditional deterministic reliability constraints. In Chapter 2 we use large deviations techniques to study the probability of current and temperature overloads in transmission lines assuming a stochastic differential equations model for random power injections and a small-noise regime. We analytically characterize the set of admissible power injections (referred to as capacity regions) such that the probability of overloading of any line over a given time interval stays below a fixed target, and prove convexity properties that make them amenable to be used within existing planning methods like Optimal Power Flow (OPF). The approach models the stochastic behavior at the process-level and takes into account the transient relationship between line current and line temperatures,

leading to capacity gains compared to static approaches. In Chapter 3 we develop explicit upper bounds for line failure probabilities based on concentration inequalities techniques which, in contrast to large deviations methods, do not assume a small-noise regime and are guaranteed to be conservative, leading to safe operational capacity regions that are convex and polyhedral.

Chapters 4 and 5 are devoted to understanding microscopic and macroscopic features of blackouts. Chapter 4 focuses on the microscopic viewpoint and models power grids as complex networks where failures can emerge endogenously as a result of the interplay between noisy correlated power inputs at the nodes, the network structure, and power flow physics. Using large deviations techniques we rank transmission lines according to their failure probability, we explicitly determine the most likely configuration of power inputs leading to line failures and predict how subsequent failures will propagate in the network. The results are mathematically exact in a small-noise regime and are validated in a realistic setting using data for the German transmission grid. In Chapter 5 we analyze cascading failures from a macroscopic perspective, and provide a causal explanation for the well-known fact that the probability distribution of blackout sizes is scale-free. We model power grids as graphs with heavy-tailed sinks, which represent demand from cities, and study cascading failures on such graphs. Combining the physics of power flows with rare event analysis for heavy-tailed distributions, our research links the scale-free nature of blackout sizes to the scale-free nature of city sizes, and is validated using synthetic networks and the German grid. Our approach differs from traditional explanations in that it does not relate scale-free phenomena to the scale-free nature of the network topology, and suggest new ways of approaching such phenomena in other transport networks.

Chapters 6 and 7 focus on understanding the impact of uncertainty on energy prices. In Chapter 6 we develop a statistical learning methodology to recover the energy market's structure and predict Locational Marginal Prices (LMP) from a decentralized perspective, i.e. by using only publicly available data. In particular, we exploit the mathematical properties of the supply-demand matching process to characterize LMPs as dual variables of the OPF and use machine learning and convex optimization techniques to infer crucial information about the congestion status of transmission lines, and predict prices based on system load and grid-wide generation type mix forecasts. Finally, Chapter 7 focuses on predicting large spikes in LMPs, which are difficult to forecast under a decentralized perspective. By taking the viewpoint of the grid operator (hence assuming full knowledge of the grid parameters), we look at LMPs as deterministic functions of the input stochastic processes and use techniques from large deviations theory to study the occurrences of extreme price spikes caused by unusual renewable generation profiles.

About the author

Tommaso Nesti was born in Pisa, Italy, on 06-12-1988. After finishing his secondary education in 2007 at "Liceo Scientifico XXV Aprile" in Pontedera, Italy, he started his studies in Mathematics at University of Pisa, Italy, from which he received his Bachelor's degree in 2012. He then pursued a Master's degree in Mathematics at the same university, and obtained his degree with highest honors (cum laude) in 2015. During his Master's, Tommaso completed an internship at BCAM (Bilbao, Spain) under the supervision of Jonatha Anselmi in 2013, where he conducted research on optimal scheduling for parallel queuing networks.

From October 2015 he started a PhD project within the Stochastics group at Centrum Wiskunde & Informatica (CWI) in Amsterdam, Netherlands, under the supervision of Bert Zwart and Alessandro Zocca. His PhD research focuses on stochastic analysis of energy networks under uncertainty, with a special interests for the analysis of rare events, such as cascading failures and blackouts, and for energy prices forecasting. The results of his research are presented in this dissertation. During his employment at CWI, Tommaso was invited to visit the Simons Institute for the Theory of Computing in Berkeley, CA, and Isaac Newton Institute for Mathematical Sciences in Cambridge, UK. Both visits were aimed at establishing research collaborations. He also completed two internships at Google Inc, Sunnyvale (CA), in 2017 and 2018 respectively, where under the supervision of Ana Radovanovic. he worked as a Data Scientist on topics related to energy prices in wholesale energy markets. He attended several workshops and conferences in the Netherlands, as well as international schools and conferences, among which Stochastic Networks 2016 in San Diego (CA) and INFORMS 2017 in Houston (TX). Tommaso will defend his PhD thesis at Eindhoven University of Technology on March 30, 2020.